On Convergence Rates Equivalency and Sampling Strategies in Functional Deconvolution Models

Marianna Pensky,
Department of Mathematics, University of Central Florida,
Orlando, FL 32816-1364, USA.

Theofanis Sapatinas,
Department of Mathematics and Statistics, University of Cyprus,
P.O. Box 20537, CY 1678 Nicosia, Cyprus.

Abstract

Using the asymptotical minimax framework, we examine convergence rates equivalency between a continuous functional deconvolution model and its real-life discrete counterpart, over a wide range of Besov balls and for the L^2-risk. For this purpose, all possible models are divided into three groups. For the models in the first group, which we call *uniform*, the convergence rates in the discrete and the continuous models coincide no matter what the sampling scheme is chosen and, hence, the replacement of the discrete model by its continuous counterpart is legitimate. For the models in the second group, to which we refer as *regular*, one can point out the best sampling strategy in the discrete model, but not every sampling scheme leads to the same convergence rates: there are at least two sampling schemes which deliver different convergence rates in the discrete model (i.e., at least one of the discrete models leads to convergence rates that are different from the convergence rates in the continuous model). The third group consists of models for which, in general, it is impossible to devise the best sampling strategy; we call these models *irregular*.

We formulate the conditions when each of these situations takes place. In the regular case, not only we point out the number and the selection of sampling points which deliver the fastest convergence rates in the discrete model but also investigate when, in the case of an arbitrary sampling scheme, the convergence rates in the continuous model coincide or do not coincide with the convergence rates in the discrete model. We also study what happens if one chooses a uniform, or a more general pseudo-uniform, sampling scheme which can be viewed as an intuitive replacement of the continuous model. Finally, as a representative of the irregular case, we study functional deconvolution with a box-car like blurring function since this model has a number of important applications. All theoretical results presented in the paper are illustrated by numerous examples many of which are motivated directly by a multitude of inverse problems in mathematical physics where one needs to recover initial or boundary conditions on the basis of observations from a noisy solution of a partial differential equation. The theoretical performance of the suggested estimator in the multichannel deconvolution model with a box-car like blurring function is also supplemented by a limited simulation study and compared to an estimator available in the current literature. The paper concludes that in both regular and irregular cases one should be extremely careful when replacing a discrete functional deconvolution model by its continuous counterpart.

Keywords: adaptivity; Besov spaces; deconvolution; Fourier analysis; Meyer wavelets; minimax estimators; multichannel deconvolution; thresholding; wavelet analysis.

AMS (2000) Subject Classification: Primary 62G05; Secondary 62G08

1 Introduction

We consider the estimation problem of the unknown response function $f(\cdot)$ based on observations from the following noisy convolutions

$$y(u,t) = f * g(u,t) + \frac{1}{\sqrt{n}}\, z(u,t), \quad u \in U, \ \ t \in T, \tag{1.1}$$

where $U = [a,b]$, $-\infty < a \leq b < \infty$, and $T = [0,1]$. Here, $z(u,t)$ is assumed to be a two-dimensional Gaussian white noise, i.e., a generalized two-dimensional Gaussian field with covariance function $\mathbb{E}[z(u_1,t_1)z(u_2,t_2)] = \delta(u_1-u_2)\delta(t_1-t_2)$, where $\delta(\cdot)$ denotes the Dirac δ-function, and

$$f * g(u,t) = \int_T f(x)g(u,t-x)dx,$$

with the blurring (or kernel) function $g(\cdot,\cdot)$ also assumed to be known.

The model (1.1) has been recently introduced by Pensky & Sapatinas (2009) and can be viewed as a *functional deconvolution* model. If $a = b$, it reduces to the standard deconvolution model which attracted attention of a number of researchers, e.g., Donoho (1995), Abramovich & Silverman (1998), Kalifa & Mallat (2003), Johnstone *et.al.* (2004), Donoho & Raimondo (2004), Johnstone & Raimondo (2004), Neelamani, Choi & Baraniuk (2004), Kerkyacharian *et.al.* (2007), Cavalier & Raimondo (2007) and Chesneau (2008).

The functional deconvolution model (1.1) can be viewed as a generalization of a multitude of inverse problems in mathematical physics where one needs to recover initial or boundary conditions on the basis of observations of a noisy solution of a partial differential equation. Lattes & Lions (1967) initiated research in the problem of recovering the initial condition for parabolic equations based on observations in a fixed-time strip, while this problem and the problem of recovering the boundary condition for elliptic equations based on observations in an internal domain were studied in Golubev & Khasminskii (1999); the latter problem was also discussed in Golubev (2004). These and other specific models in mathematical physics were discussed in detail in Pensky & Sapatinas (2009).

However, model (1.1) is just an idealization of a real-life situation. One can make observations only at particular points (u_l, t_i), $l = 1,2,\ldots,M$, $i = 1,2,\ldots,N$, so that the actual problem can be formulated as follows: recover the unknown response function $f(\cdot)$ from observations $y(u_l,t_i)$, where

$$y(u_l,t_i) = \int_T f(x)g(u_l,t_i-x)dx + \varepsilon_{li}, \quad u_l \in U, \ \ t_i = i/N, \tag{1.2}$$

with ε_{li} being standard Gaussian random variables, independent for different l and i. Model (1.2) can be viewed as a *discrete* version of the *continuous* functional deconvolution model (1.1).

It is well documented in the literature that asymptotic equivalence between discrete and continuous models holds in some nonparametric models. In particular, Brown & Low (1996) and Brown, Cai, Low & Zhang (2002) in the univariate case, and Reiss (2008) in the multivariate case, established, under some restrictions, asymptotic equivalence (in Le Cam sense) between nonparametric regression and Gaussian white noise models. Although, to the best of our knowledge, such an asymptotic equivalence between continuous and discrete models, in the functional deconvolution setting, has not yet been explored, it has been documented in the literature a convergence rate equivalency, in the asymptotical minimax sense, between standard continuous and discrete deconvolution models, i.e., when $a = b$, $M = 1$ and $N = n$ in (1.1) and (1.2), over a wide range of Besov balls and for the L^r-risks, $1 \leq r < \infty$, e.g., Chesneau (2008), Pensky & Sapatinas (2009) and Petsa & Sapatinas (2009).

For the above reason, and using the asymptotical minimax framework, one may attempt to study the continuous functional deconvolution model (1.1) instead of its discrete counterpart (1.2), assuming that the convergence rates between these models coincide. However, in this case, this equivalence has only a limited scope. Indeed, Pensky & Sapatinas (2009) only touched upon the issue, showing that, under very restrictive conditions, a convergence rate equivalence between the continuous functional deconvolution model (1.1) and its discrete counterpart (1.2) models holds when $n = NM$, over a wide range of Besov balls and for the L^2-risk. Nevertheless, in majority of practical situations, these conditions are violated and it remains to be seen how legitimate the replacement of the real life model (1.2) by its idealization (1.1) is, even in the case of inverse problems in mathematical physics, presented in Pensky & Sapatinas (2009). In fact, in many situations, the convergence rates in the two models depend on the choice of M and the selection of sampling points $u_1, u_2, \ldots, u_M$ and may coincide with the convergence rates in the continuous model for one selection and be different for another. Also, from a practical point of view, the objective is not to find M and $u_1, u_2, \ldots, u_M$ which make the two models equivalent from the convergence rate viewpoint, but rather to point out M and $u_1, u_2, \ldots, u_M$ which deliver the fastest possible convergence rates in the real life model (1.2). Note that the discrete model (1.2) can also be viewed as a multichannel deconvolution model where the number of channels $M = M_n$ is fixed or, possibly, $M_n \to \infty$ as the sample size $n \to \infty$; the case when $M \geq 2$ (finite) was considered in, e.g., Casey & Walnut (1994) and De Canditiis & Pensky (2004, 2006). Hence, if the kernel $g(\cdot, \cdot)$ is fixed, the choice of M and the selection of sampling points $u_1, u_2, \ldots, u_M$ which provide the fastest convergence rates is of extreme importance in signal processing.

Using the asymptotical minimax framework, our objective is to evaluate how legitimate it is to replace the real-life discrete model (1.2) by its continuous counterpart (1.1). For this purpose, we shall divide all possible models into three groups. For the models in the first group, which we call *uniform*, the convergence rates in discrete and continuous functional deconvolution models coincide no matter what the sampling scheme is chosen and, hence, the replacement of the discrete model by its continuous counterpart is legitimate. For the models in the second group, to which we refer as *regular*, one can point out the best sampling strategy in the discrete model (i.e., the strategy which leads to the fastest convergence rate), but not every sampling scheme leads to the same convergence rates: there are at least two sampling schemes which deliver different convergence rates in the discrete model (i.e., at least one of the discrete models leads to convergence rates that are different from the convergence rates in the continuous functional deconvolution model). The third group consists of models for which, in general, it is impossible to devise the best sampling strategy; we call these models *irregular*.

We formulate the conditions when each of these situations takes place. In the regular case, not only we point out the choice of M and the selection of sampling points $u_1, u_2, \ldots, u_M$ which deliver the fastest convergence rates in the discrete model but also investigate when, in the case of an arbitrary sampling scheme, the convergence rates in the continuous functional deconvolution model coincide or do not coincide with the convergence rates of its discrete counterpart. We also study what happens if one chooses a uniform, or a more general pseudo-uniform, sampling scheme which can be viewed as an intuitive replacement of the continuous model. Finally, as a representative of the irregular case, we study functional deconvolution with a box-car like kernel since this model has a number of important applications. All theoretical results presented are illustrated by numerous examples many of which are motivated directly by a multitude of inverse problems in mathematical physics where one needs to recover initial or boundary conditions on the basis of observations from a noisy solution of a partial differential equation.

As in Pensky & Sapatinas (2009), we consider functional deconvolution in a periodic setting, i.e., we assume that $f(\cdot)$ and, for fixed $u \in U$, $g(u, \cdot)$ are periodic functions with period on the

unit interval T. Note that the periodicity assumption appears naturally in the above mentioned special models which (1.1) and (1.2) generalize, and allows one to explore ideas considered in the above cited papers to the proposed functional deconvolution framework. Moreover, not only for theoretical reasons but also for practical convenience (see Johnstone *et.al.* (2004), Sections 2.3, 3.1–3.2), we use band-limited wavelet bases, and in particular the periodized Meyer wavelet basis for which fast algorithms exist (see Kolaczyk (1994) and Donoho & Raimondo (2004)). In order to also allow inhomogeneous functions $f(\cdot)$ into our study, we consider a wide range of Besov balls, as it is common in the wavelet literature and, for simplicity, we work with the L^2-risk only. However, the results of this paper can be extended to a more general class of L^r-risks, $1 \le r < \infty$, using the unconditionality and Temlyakov properties of Meyer wavelets, e.g., Johnstone *et.al.* (2004) and Petsa & Sapatinas (2009).

The rest of the paper is organized as follows. In Section 2, we describe the construction of wavelet estimators of $f(\cdot)$ derived by Pensky & Sapatinas (2009) both for the continuous functional deconvolution model (1.1) and its discrete counterpart (1.2). In Section 3, for the continuous model and for a discrete model with any particular design of sampling points, using the asymptotical minimax framework, we provide lower bounds for the L^2-risk over a wide range of Besov balls and show that those bounds are attained by the wavelet estimators constructed in Section 2. Section 4 is devoted to the discussion of the interplay between continuous and discrete functional deconvolution models. First, Section 4.1 formulates the necessary and sufficient conditions for the convergence rates in continuous and discrete functional deconvolution models to coincide and to be independent of the choice of M and the selection of points $u_1, u_2, \ldots, u_M$. Then, Section 4.2 provides examples where these conditions do or do not take place, and Section 4.3 sorts all possible situations into the uniform, regular and irregular cases. Section 5 studies the regular case. In particular, Section 5.1 designs the best possible sampling strategy. Section 5.2 provides some motivating examples. Section 5.3 investigates the relation between the L^2-risks in the continuous and the discrete models under an arbitrary sampling scheme and formulates conditions when the convergence rates do or do not coincide. Section 5.4 formulates sufficient conditions when the convergence rates in both models coincide under a pseudo-uniform sampling scheme in the discrete model. Section 5.5 provides a variety of examples where the convergence rates coincide or differ depending on what sampling scheme is employed. Section 6 explores the interplay between continuous and discrete functional deconvolution models in the case of a box-car like blurring function. Section 7 supplements the theory with a limited simulation study in the case of a box-car like blurring function and compares performance of the suggested estimator to the estimator proposed by De Canditiis & Pensky (2006). Concluding remarks are given in Section 8. Finally, Section 9 (Appendix) provides the proofs of the theoretical results obtained in the previous sections.

In the rest of the paper, the continuous functional deconvolution model (1.1) is referred to as the "continuous model" and its discrete counterpart (1.2) is referred to as the "discrete model".

2 Wavelet estimators

For both the continuous model and the discrete model, we use the wavelet estimator derived in Pensky & Sapatinas (2009), described as follows.

Let $\varphi^*(\cdot)$ and $\psi^*(\cdot)$ be the Meyer scaling and mother wavelet functions, respectively, in the real line (see, e.g., Meyer (1992)) and obtain a periodized version of Meyer wavelet basis as in Johnstone *et.al.* (2004), i.e., for $j \ge 0$ and $k = 0, 1, \ldots, 2^j - 1$,

$$\varphi_{jk}(x) = \sum_{i \in \mathbb{Z}} 2^{j/2} \varphi^*(2^j(x+i) - k), \quad \psi_{jk}(x) = \sum_{i \in \mathbb{Z}} 2^{j/2} \psi^*(2^j(x+i) - k), \quad x \in T.$$

Denote $\langle f, g\rangle = \int_T f(t)\overline{g(t)}dt$, the inner product in the Hilbert space $L^2(T)$. Let $e_m(t) = e^{i2\pi mt}$, $m \in \mathbb{Z}$, and, for any (primary resolution level) $j_0 \geq 0$ and any $j \geq j_0$, let $\varphi_{mj_0k} = \langle e_m, \varphi_{j_0k}\rangle$, $\psi_{mjk} = \langle e_m, \psi_{jk}\rangle$, $f_m = \langle e_m, f\rangle$ be the Fourier coefficients of $\varphi_{jk}(\cdot)$, $\psi_{jk}(\cdot)$ and $f(\cdot)$, respectively. For each $u \in U$, denote the functional Fourier coefficients by

$$y_m(u) = \langle e_m, y(u,\cdot)\rangle, \quad g_m(u) = \langle e_m, g(u,\cdot)\rangle.$$

In what follows we assume that function $g(u,t)$ is such that $g_m(u)$ are continuous functions of u for every m. (This condition is not restrictive and holds in all examples considered below.)

If we have the continuous model (1.1), then, by using properties of the Fourier transform, for each $u \in U$, we have $h_m(u) = g_m(u)f_m$ and

$$y_m(u) = g_m(u)f_m + n^{-1/2}z_m(u), \tag{2.1}$$

where $z_m(u)$ are generalized one-dimensional Gaussian processes such that $\mathbb{E}[z_{m_1}(u_1)z_{m_2}(u_2)] = \delta_{m_1,m_2}\delta(u_1 - u_2)$, where δ_{m_1,m_2} is Kronecker's delta. If we have the discrete model (1.2), then, by using properties of the discrete Fourier transform, for each $l = 1, 2, \ldots, M$, (2.1) takes the form

$$y_m(u_l) = g_m(u_l)f_m + N^{-1/2}z_{ml}, \tag{2.2}$$

where z_{ml} are standard Gaussian random variables, independent for different m and l.

Estimate the Fourier coefficients f_m of $f(\cdot)$ by

$$\widehat{f}_m = \left(\int_a^b \overline{g_m(u)}y_m(u)du\right)\Big/\left(\int_a^b |g_m(u)|^2du\right) \text{ in the continuous model,} \tag{2.3}$$

$$\widehat{f}_m = \left(\sum_{l=1}^{M} \overline{g_m(u_l)}y_m(u_l)\right)\Big/\left(\sum_{l=1}^{M} |g_m(u_l)|^2\right) \text{ in the discrete model.} \tag{2.4}$$

Here, we adopt the convention that when $a = b$, $\widehat{f}_m$ takes the form $\widehat{f}_m = (\overline{g_m(a)}y_m(a)/|g_m(a)|^2$ and somewhat abuse notation using f_m for both functional Fourier coefficients and their discrete counterparts.

Note that, using the periodized Meyer wavelet basis described above and for any $j_0 \geq 0$, any (periodic) $f(\cdot) \in L^2(T)$ can be expanded as

$$f(t) = \sum_{k=0}^{2^{j_0}-1} a_{j_0k}\varphi_{j_0k}(t) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} b_{jk}\psi_{jk}(t), \quad t \in T. \tag{2.5}$$

Furthermore, by Plancherel's formula, the scaling coefficients, $a_{j_0k} = \langle f, \varphi_{j_0k}\rangle$, and the wavelet coefficients, $b_{jk} = \langle f, \psi_{jk}\rangle$, of $f(\cdot)$ can be represented as

$$a_{j_0k} = \sum_{m\in C_{j_0}} f_m\overline{\varphi_{mj_0k}}, \quad b_{jk} = \sum_{m\in C_j} f_m\overline{\psi_{mjk}}, \tag{2.6}$$

where $C_{j_0} = \{m : \varphi_{mj_0k} \neq 0\}$ and, for any $j \geq j_0$, $C_j = \{m : \psi_{mjk} \neq 0\}$. We estimate a_{j_0k} and b_{jk} by substituting f_m in (2.6) with (2.3) or (2.4), i.e.,

$$\widehat{a}_{j_0k} = \sum_{m\in C_{j_0}} \widehat{f}_m\overline{\varphi_{mj_0k}}, \quad \widehat{b}_{jk} = \sum_{m\in C_j} \widehat{f}_m\overline{\psi_{mjk}}. \tag{2.7}$$

We now construct a (block thresholding) wavelet estimator of $f(\cdot)$, suggested by Pensky & Sapatinas (2009). For this purpose, we divide the wavelet coefficients at each resolution level into blocks of length $\ln n$. Let A_j and U_{jr} be the following sets of indices $A_j = \{r \mid r = 1, 2, \ldots, 2^j/\ln n\}$, $U_{jr} = \{k \mid k = 0, 1, \ldots, 2^j - 1;\ (r-1)\ln n \le k \le r \ln n - 1\}$. Denote

$$B_{jr} = \sum_{k \in U_{jr}} b_{jk}^2, \quad \widehat{B}_{jr} = \sum_{k \in U_{jr}} \widehat{b}_{jk}^2. \tag{2.8}$$

Finally, for any $j_0 \ge 0$, $f(\cdot)$ is constructed as

$$\hat{f}_n(t) = \sum_{k=0}^{2^{j_0}-1} \widehat{a}_{j_0 k} \varphi_{j_0 k}(t) + \sum_{j=j_0}^{J-1} \sum_{r \in A_j} \sum_{k \in U_{jr}} \widehat{b}_{jk} \mathbb{I}(|\widehat{B}_{jr}| \ge \lambda_j)\, \psi_{jk}(t), \quad t \in T, \tag{2.9}$$

where $\mathbb{I}(A)$ is the indicator function of the set A, and the resolution levels j_0 and J and the thresholds λ_j will be defined in Section 3.2.

In what follows, the symbol C is used for a generic positive constant, independent of n, while the symbol K is used for a generic positive constant, independent of m, n, M and $u_1, u_2, \ldots, u_M$, which either of them may take different values at different places.

3 Minimax lower and upper bounds for the L^2-risk over Besov balls

Among the various characterizations of Besov spaces for periodic functions defined on $L^p(T)$ in terms of wavelet bases, we recall that for an r-regular multiresolution analysis with $0 < s < r$ and for a Besov ball, $B_{p,q}^s(A) = \{f(\cdot) \in L^p(T) :\ f \in B_{p,q}^s,\ \|f\|_{B_{p,q}^s} \le A\}$, of radius $A > 0$ with $1 \le p, q \le \infty$, one has that, with $s' = s + 1/2 - 1/p$,

$$B_{p,q}^s(A) = \left\{ f(\cdot) \in L^p(T) : \left(\sum_{k=0}^{2^{j_0}-1} |a_{j_0 k}|^p \right)^{\frac{1}{p}} + \left(\sum_{j=j_0}^{\infty} 2^{j s' q} \left(\sum_{k=0}^{2^j-1} |b_{jk}|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}} \le A \right\}, \tag{3.1}$$

with respective sum(s) replaced by maximum if $p = \infty$ or $q = \infty$ (see, e.g., Johnstone *et.al.* (2004), Section 2.4). (Note that, for the Meyer wavelet basis, considered in Section 2, $r = \infty$.)

We construct below asymptotical minimax lower bounds for the L^2-risk based on observations from either the continuous model or the discrete model. For this purpose, we define the corresponding minimax L^2-risks over the set Ω as

$$R_n^c(\Omega) = \inf_{\tilde{f}_n^c} \sup_{f \in \Omega} \mathbb{E}\|\tilde{f}_n^c - f\|^2, \tag{3.2}$$

$$R_n^d(\Omega, \underline{u}, M) = \inf_{\tilde{f}_n^d} \sup_{f \in \Omega} \mathbb{E}\|\tilde{f}_n^d - f\|^2, \tag{3.3}$$

$$R_n^d(\Omega) = \inf_{\underline{u}, M} R_n^d(\Omega, \underline{u}, M), \tag{3.4}$$

where the infimum in (3.2) is taken over all possible estimators (i.e., measurable functions) $\tilde{f}_n^c(\cdot)$ of $f(\cdot)$ from the continuous model, the infimum in (3.4) is taken over all possible estimators $\tilde{f}_n^d(\cdot)$ of $f(\cdot)$ from the discrete model, based on a sample at M points $\underline{u} = (u_1, u_2 \ldots, u_M)$, and the infimum in (3.4) is evaluated over all possible estimators $\tilde{f}_n^d(\cdot)$ of $f(\cdot)$ and the choices of M and $\underline{u}$. Note that, since the asymptotical minimax convergence rates for the L^2-risk in the discrete model

depends on M and $\underline{u}$ if these quantities are fixed, we are interested in the selection of M and $\underline{u}$, minimizing the asymptotical minimax convergence rates for the L^2-risk.

Denote $s^* = s + 1/2 - 1/p'$, $\quad p' = \min(p, 2)$, and, for $\kappa = 1, 2$, define

$$\tau_\kappa^c(m) = \int_a^b |g_m(u)|^{2\kappa}\, du \quad \text{and} \quad \tau_\kappa^d(m, \underline{u}, M) = \frac{1}{M} \sum_{l=1}^{M} |g_m(u_l)|^{2\kappa}, \tag{3.5}$$

where $\tau_1^c(m) = |g_m(a)|^2$ when $a = b$.

Pensky & Sapatinas (2009) constructed asymptotical minimax lower and upper bounds for the L^2-risk for the continuous model. The corresponding bounds for the discrete model were obtained under the very restrictive conditions that the upper and the lower bounds on $\tau_1^d(m, \underline{u}, M)$ do not depend on n, M and $\underline{u}$. Below we shall need asymptotic lower and upper bounds for the L^2-risk in the case of much more general expressions for $\tau_1^c(m)$ and $\tau_1^d(m, \underline{u}, M)$, than in Pensky & Sapatinas (2009).

3.1 Minimax lower bounds: particular choice of sampling points

Let, with some abuse of notation, $\tau_1(m) = \tau_1^c(m)$, $R_n^*(B_{p,q}^s(A)) = R_n^c(B_{p,q}^s(A))$, in the continuous model, and $\tau_1(m) = \tau_1^d(m, \underline{u}, M)$, $R_n^*(B_{p,q}^s(A)) = R_n^d(B_{p,q}^s(A), \underline{u}, M)$, in the discrete model.

Assume that for some constants $\nu \in \mathbb{R}$, $\lambda \in \mathbb{R}$, $\alpha \geq 0$ and $\beta > 0$, independent of m and n, and for some sequence $\varepsilon_n > 0$, independent of m,

$$\tau_1(m) \leq K\varepsilon_n\, |m|^{-2\nu} (\ln|m|)^{-\lambda} \exp\big(-\alpha|m|^\beta\big), \quad \nu > 0 \quad \text{if} \quad \alpha = 0. \tag{3.6}$$

Denote $n^* = n\varepsilon_n$ and assume that the sequence ε_n is such that

$$n^* = n\varepsilon_n \to \infty \quad \text{as} \quad n \to \infty. \tag{3.7}$$

Then, the following statement is true.

Theorem 1 *Let $\{\phi_{j_0,k}(\cdot), \psi_{j,k}(\cdot)\}$ be the periodic Meyer wavelet basis discussed in Section 2. Let $s > \max(0, 1/p - 1/2)$, $1 \leq p \leq \infty$, $1 \leq q \leq \infty$ and $A > 0$. Let assumptions (3.6) and (3.7) hold. Then, as $n \to \infty$,*

$$R_n^*(B_{p,q}^s(A)) \geq \begin{cases} C(n^*)^{-\frac{2s}{2s+2\nu+1}}\, (\ln n^*)^{\frac{2s\lambda}{2s+2\nu+1}}, & \text{if} \quad \alpha = 0,\ \nu(2-p) < ps^*, \\ C\left(\frac{\ln n^*}{n^*}\right)^{\frac{2s^*}{2s^*+2\nu}} (\ln n^*)^{\frac{2s^*\lambda}{2s^*+2\nu}}, & \text{if} \quad \alpha = 0,\ \nu(2-p) \geq ps^*, \\ C(\ln n^*)^{-\frac{2s^*}{\beta}}, & \text{if} \quad \alpha > 0. \end{cases} \tag{3.8}$$

3.2 Minimax upper bounds: particular choice of sampling points

Assume that for the constants $\nu \in \mathbb{R}$, $\lambda \in \mathbb{R}$, $\alpha \geq 0$ and $\beta > 0$ and the sequence $\varepsilon_n > 0$ in (3.6)

$$\tau_1(m) \geq K\varepsilon_n |m|^{-2\nu} (\ln|m|)^{-\lambda} \exp\big(-\alpha|m|^\beta\big), \quad \nu > 0 \quad \text{if} \quad \alpha = 0. \tag{3.9}$$

Let $\hat{f}_n(\cdot)$ be the wavelet estimator defined by (2.9). Let, as before, $n^* = n\varepsilon_n$ satisfy condition (3.7), and assume that in the case of $\alpha = 0$ in (3.9) the sequence ε_n is such that

$$-h_1 \ln n \leq \ln(1/\varepsilon_n) \leq (1 - h_2) \ln n \tag{3.10}$$

for some constants $h_1 > 0$ and $h_2 \in (0,1)$. Observe that condition (3.10) implies (3.7) and that $\ln n^* \asymp \ln n$ as $n \to \infty$. Here, and in what follows, $u(n) \asymp v(n)$ means that there exist constants $C_1 > 0$ and $C_2 > 0$, independent of n, such that $0 < C_1 v(n) \le u(n) \le C_2 v(n) < \infty$ for n large enough.

Choose j_0 and J such that

$$2^{j_0} = \ln(n^*), \quad 2^J = (n^*)^{\frac{1}{2\nu+1}}, \quad \text{if} \quad \alpha = 0, \tag{3.11}$$

$$2^{j_0} = \frac{3}{8\pi}\left(\frac{\ln(n^*)}{2\alpha}\right)^{\frac{1}{\beta}}, \quad 2^J = 2^{j_0}, \quad \text{if} \quad \alpha > 0. \tag{3.12}$$

(Since $j_0 > J-1$ when $\alpha > 0$, the estimator (2.9) only consists of the first (linear) part and, hence, λ_j does not need to be selected in this case.) Set, for some constant $\mu > 0$, large enough,

$$\lambda_j = \mu^2\, (n^*)^{-1} \ln(n^*)\, 2^{2\nu j} j^{\lambda}, \quad \text{if} \quad \alpha = 0. \tag{3.13}$$

Note that the choices of j_0, J and λ_j are independent of the parameters, s, p, q and A of the Besov ball $B^s_{p,q}(A)$; hence, the estimator (2.9) is adaptive with respect to these parameters.

Set $(x)_+ = \max(0,x)$, and define

$$\varrho = \begin{cases} \frac{(2\nu+1)(2-p)_+}{p(2s+2\nu+1)}, & \text{if} \quad \nu(2-p) < ps^*, \\ \frac{(q-p)_+}{q}, & \text{if} \quad \nu(2-p) = ps^*, \\ 0, & \text{if} \quad \nu(2-p) > ps^*. \end{cases} \tag{3.14}$$

For any $j \ge j_0$, let $|C_j|$ be the cardinality of the set C_j; note that, for Meyer wavelets, $|C_j| = 4\pi 2^j$ (see, e.g., Johnstone *et.al.* (2004)). Let also

$$\Delta_\kappa(j) = |C_j|^{-1} \sum_{m \in C_j} \tau_\kappa(m)\, [\tau_1(m)]^{-2\kappa}, \quad \kappa = 1, 2. \tag{3.15}$$

Direct calculations yield that under conditions (3.9) and (3.10), for some constants $c_1 > 0$ and $c_2 > 0$, independent of n, one has

$$\Delta_1(j) \le \begin{cases} c_1\ (\varepsilon_n)^{-1}\ 2^{2\nu j} j^{\lambda}, & \text{if} \quad \alpha = 0, \\ c_2\ (\varepsilon_n)^{-1}\ 2^{2\nu j} j^{\lambda}\ \exp\left\{\alpha\left(\frac{8\pi}{3}\right)^{\beta} 2^{j\beta}\right\}, & \text{if} \quad \alpha > 0. \end{cases} \tag{3.16}$$

The proof of the minimax upper bounds for the L^2-risk is based on the following two lemmas.

Lemma 1 *Let assumption (3.9) hold, and let the estimators $\widehat{a}_{j_0 k}$ and $\widehat{b}_{jk}$ of the scaling and wavelet coefficients $a_{j_0 k}$ and b_{jk}, respectively, be given by the formula (2.7) with $\widehat{f}_m$ defined by (2.3) in the continuous model and by (2.4) in the discrete model. Then, for all $j \ge j_0$,*

$$\mathbb{E}|\widehat{a}_{j_0 k} - a_{j_0 k}|^2 \le C n^{-1} \Delta_1(j_0), \quad \mathbb{E}|\widehat{b}_{jk} - b_{jk}|^2 \le C n^{-1} \Delta_1(j). \tag{3.17}$$

If $\alpha = 0$ and assumption (3.10) holds, then, for any $j \ge j_0$, one has

$$\mathbb{E}|\widehat{b}_{jk} - b_{jk}|^4 \le C n (\ln n)^{3\lambda}\, (n^*)^{-\frac{3}{2\nu+1}}. \tag{3.18}$$

Lemma 2 *Let the estimators* $\widehat{b}_{jk}$ *of the wavelet coefficients* b_{jk} *be given by the formula (2.7) with* $\widehat{f}_m$ *defined by (2.3) in the continuous model and by (2.4) in the discrete model. Let assumptions (3.9) (if* $\alpha = 0$*) and (3.10) hold. If*

$$\mu \geq 2\sqrt{c_1}(\sqrt{6}+1)/\sqrt{h_2}, \tag{3.19}$$

where c_1 *and* h_2 *are defined in (3.16) and (3.10), respectively, then, for all* $j \geq j_0$*,*

$$\mathbb{P}\left(\sum_{k \in U_{jr}} |\widehat{b}_{jk} - b_{jk}|^2 \geq (4n^*)^{-1}\ \mu^2\, 2^{2\nu j}\, j^{\lambda}\, \ln(n^*) \right) \leq n^{-3}. \tag{3.20}$$

Then, the following statement is true.

Theorem 2 *Let* $\hat{f}_n(\cdot)$ *be the wavelet estimator defined by (2.9), with* j_0 *and* J *given by (3.11) (if* $\alpha = 0$*) or (3.12) (if* $\alpha > 0$*) and* μ *satisfying (3.19). Let* $s > 1/p'$, $1 \leq p \leq \infty$, $1 \leq q \leq \infty$ *and* $A > 0$*. Then, under the assumptions (3.9) and (3.7) if* $\alpha > 0$*, or (3.9) and (3.10) if* $\alpha = 0$*, as* $n \to \infty$*,*

$$\sup_{f \in B^s_{p,q}(A)} \mathbb{E}\|\hat{f}_n - f\|^2 \leq \begin{cases} C(n^*)^{-\frac{2s}{2s+2\nu+1}}\ (\ln n)^{\varrho + \frac{2s\lambda}{2s+2\nu+1}}, & \text{if} \quad \alpha = 0,\ \nu(2-p) < ps^*, \\ C\left(\frac{\ln n}{n^*}\right)^{\frac{2s^*}{2s^*+2\nu}}\ (\ln n)^{\varrho + \frac{2s^*\lambda}{2s^*+2\nu}}, & \text{if} \quad \alpha = 0,\ \nu(2-p) \geq ps^*, \\ C(\ln(n^*))^{-\frac{2s^*}{\beta}}, & \text{if} \quad \alpha > 0. \end{cases} \tag{3.21}$$

Remark 1 Note that in the continuous model, one can write a lower bound for $\tau_1(m)$ in (3.6) and an upper bound for $\tau_1(m)$ in (3.9) with $\varepsilon_n = 1$, so that $n^* = n$ in (3.8) and (3.21). However, in the discrete model this may not be possible. Theorems 1 and 2 allow to account for the dependence of $\tau_1(m)$ on n in the case of the discrete model as well as for an extra logarithmic factor in the expression of $\tau_1(m)$ which often appears in the case of the continuous model.

Remark 2 Note that Theorems 1 and 2 can be applied even if the values of ν, λ, α and β in assumptions (3.6) and (3.9) are different, that may also depend on M and $\underline{u}$. Then, Theorem 1 provides asymptotical minimax lower bounds for the L^2-risk while Theorem 2 provides the corresponding upper bounds. If, in the continuous model or in the discrete model with some particular choice of M and sampling points $\underline{u}$, the values of ν, λ, α and β and the functions ε_n in conditions (3.6) and (3.9) coincide, then Theorems 1 and 2 imply that the estimator $\hat{f}_n(\cdot)$ defined by (2.9) is asymptotically optimal (in the minimax sense), or near-optimal within a logarithmic factor, over a wide range of Besov balls $B^s_{p,q}(A)$. Therefore, in the rest of the paper, when we talk about *convergence rates* we refer to the asymptotical minimax lower bounds for the L^2-risk which are attainable, up to at most a logarithmic factor, according to Theorems 1 and 2.

4 The interplay between continuous and discrete models: uniform, regular and irregular cases

The convergence rates in the discrete model depend on two aspects: the total number of observations $n = NM$ and the behavior of $\tau_1^d(m, \underline{u}, M)$. In the continuous model, the values of $\tau_1^c(m)$ are fixed; they depend on m only and, hence, conditions (3.6) and (3.9) can be easily verified. However, this is no longer true in the discrete model; in this case, the values of $\tau_1^d(m, \underline{u}, M)$ may

depend on the choice of M and the selection of points $\underline{u}$. If we require the values of $\tau_1^d(m, \underline{u}, M)$ to be independent of the choice of M and the selection of points $\underline{u}$, then the convergence rates in the discrete and the continuous models coincide and are independent of the selection of points $\underline{u}$. Moreover, in this case, the wavelet estimator (2.9) is asymptotically optimal (in the minimax sense) no matter what the choice of M is. It is quite possible, however, that in the discrete model, conditions (3.6) and (3.9) both hold but with different values of ν, λ, α and β for different choices of M and $\underline{u}$. In this case, the asymptotical minimax upper bounds for the L^2-risk in the discrete model may not coincide with the convergence rates in the continuous model, at least for some sampling schemes.

4.1 Necessary and sufficient conditions for convergence rates equivalency between continuous and discrete models

Assume that there exist points $u_*, u^* \in [a, b]$, independent of m, such that, for any $u \in [a, b]$,

$$|g_m(u)| \le K|g_m(u^*)| \text{ and } |g_m(u)| \ge K|g_m(u_*)|. \tag{4.1}$$

In this case, $(b-a)K^2|g_m(u_*)|^2 \le \tau_1^c(m) \le (b-a)K^2|g_m(u^*)|^2$ and $K^2|g_m(u_*)|^2 \le \tau_1^d(m, \underline{u}, M) \le K^2|g_m(u^*)|^2$ for any M and $\underline{u}$. Note that, based on the assumption on the blurring function $g(\cdot, \cdot)$ made in Section 2, points u_* and u^* satisfying condition (4.1) always exist; however, they are not necessarily independent of m.

Here, and in what follows, $u_m \asymp v_m$ means that there exist constants $C_1 > 0$ and $C_2 > 0$, independent of m, such that $0 < C_1 v_m \le u_m \le C_2 v_m < \infty$ for $|m|$ large enough.

The following statement, which substantially extends Proposition 1 of Pensky & Sapatinas (2009), presents the necessary and sufficient conditions for the convergence rates in the discrete model to be independent of the choice of M and the selection of points $\underline{u}$ and, hence, to coincide with the convergence rates in the continuous model.

Theorem 3 *Let there exist constants $\nu_1 \in \mathbb{R}$, $\nu_2 \in \mathbb{R}$, $\alpha_1 \ge 0$, $\alpha_2 \ge 0$, $\beta_1 > 0$ and $\beta_2 > 0$, independent of m and n, such that*

$$|g_m(u^*)|^2 \asymp |m|^{-2\nu_1} \exp(-\alpha_1 |m|^{\beta_1}), \quad \nu_1 > 0 \quad \text{if} \quad \alpha_1 = 0, \tag{4.2}$$

$$|g_m(u_*)|^2 \asymp |m|^{-2\nu_2} \exp(-\alpha_2 |m|^{\beta_2}), \quad \nu_2 > 0 \quad \text{if} \quad \alpha_2 = 0. \tag{4.3}$$

Then, the convergence rates obtained in Theorems 1 and 2 in the discrete model are independent of the choice of M and the selection of points $\underline{u}$, and, hence, coincide with the convergence rates obtained in Theorems 1 and 2 in the continuous model, if and only if

$$\alpha_1\alpha_2 > 0 \textit{ and } \beta_1 = \beta_2 \textit{ or } \alpha_1 = \alpha_2 = 0 \textit{ and } \nu_1 = \nu_2. \tag{4.4}$$

Remark 3 Theorem 3 provides necessary and sufficient conditions for the convergence rates in the continuous and the discrete models to coincide, and to be independent of the choice of M and the selection of points $\underline{u}$. These conditions also guarantee asymptotical optimality (in the minimax sense) of the wavelet estimator (2.9), and can be viewed as some kind of uniformity conditions. Under assumptions (4.2)–(4.4), asymptotically (up to a constant factor) it makes absolutely no difference whether one samples the discrete model n times at one point, say, u_1 or, say, $\sqrt{n}$ times at $M = \sqrt{n}$ points u_l, $l = 1, 2, \ldots, M$. In other words, each sample value $y(u_l, t_i)$, $l = 1, 2, \ldots, M$, $i = 1, 2, \ldots, N$, asymptotically (up to a constant factor) gives the same amount of information and, therefore, the convergence rates are not sensitive to the choice of M and the selection of

points $\underline{u}$. On the other hand, if the conditions of Theorem 3 are violated, then the convergence rates in the discrete model depends on the choice of M and $\underline{u}$, and some recommendations on their selection should be given. Furthermore, optimality (in the minimax sense) issues become much more complex when $\tau_1^d(m; \underline{u}, M)$ is not uniformly bounded from above or below.

4.2 Some illustrative examples

Theorem 3 provides necessary and sufficient conditions for the continuous and the discrete models to be equivalent, from the viewpoint of convergence rates, no matter what the choice of M and the selection of points $\underline{u}$ are. The difficulty, however, is that many models do not satisfy those conditions. Below, we consider some illustrative examples that have recently been studied in Pensky & Sapatinas (2009).

Example 1 Estimation of the initial condition in the heat conductivity equation. Let $h(t,x)$ be a solution of the heat conductivity equation

$$\frac{\partial h(t,x)}{\partial t} = \frac{\partial^2 h(t,x)}{\partial x^2}, \quad x \in [0,1], \quad t \in [a,b], \quad a > 0,\, b < \infty,$$

with initial condition $h(0,x) = f(x)$ and periodic boundary conditions $h(t,0) = h(t,1)$ and $\partial h(t,x)/\partial x|_{x=0} = \partial h(t,x)/\partial x|_{x=1}$.

We assume that a noisy solution $y(t,x) = h(t,x) + n^{-1/2} z(t,x)$ is observed, where $z(t,x)$ is a generalized two-dimensional Gaussian field with covariance function $\mathbb{E}[z(t_1,x_1)z(t_2,x_2)] = \delta(t_1 - t_2)\delta(x_1 - x_2)$, and the goal is to recover the initial condition $f(\cdot)$ on the basis of observations $y(t,x)$. This problem was initially considered by Lattes & Lions (1967) and further studied by Golubev & Khasminskii (1999).

Then, the functional Fourier coefficients $g_m(\cdot)$ are of the form $g_m(u) = \exp(-4\pi^2 m^2 u)$, so that $u_* = b$, $u^* = a$, $|g_m(u_*)| = \exp(-4\pi^2 b m^2)$ and $|g_m(u^*)| = \exp(-4\pi^2 a m^2)$ (see Example 1 in Pensky & Sapatinas (2009)). Hence, Theorem 3 holds with $\nu_1 = \nu_2 = 0$, $\alpha_1 = 4\pi^2 b$, $\alpha_2 = 4\pi^2 a$ and $\beta_1 = \beta_2 = 2$. Therefore, the convergence rates in the continuous and the discrete models coincide, and are independent of the choice of M and the selection of points $\underline{u}$.

Example 2 Estimation of the boundary condition for the Dirichlet problem of the Laplacian on the unit circle. Let $h(x,w)$ be a solution of the Dirichlet problem of the Laplacian on a region D on the plane

$$\frac{\partial^2 h(x,w)}{\partial x^2} + \frac{\partial^2 h(x,w)}{\partial w^2} = 0, \quad (x,w) \in D \subseteq \mathbb{R}^2, \tag{4.5}$$

with a boundary ∂D and boundary condition $h(x,w)\big|_{\partial D} = F(x,w)$. Consider the situation when D is the unit circle. Then, it is advantageous to rewrite the function $h(\cdot,\cdot)$ in polar coordinates as $h(x,w) = h(u,t)$, where $u \in [0,1]$ is the polar radius and $t \in [0, 2\pi]$ is the polar angle. Then, the boundary condition can be presented as $h(1,t) = f(t)$, and $h(u,\cdot)$ and $f(\cdot)$ are periodic functions of t with period 2π.

Suppose that only a noisy version $y(u,t) = h(u,t) + n^{-1/2} z(u,t)$ is observed, where $z(u,t)$ is as in Example 1, and that observations are available only on the interior of the unit circle with $u \in [0, r_0]$, $r_0 < 1$, i.e., $a = 0$, $b = r_0 < 1$. The goal is to recover the boundary condition $f(\cdot)$ on the basis of observations $y(u,t)$. This problem was initially investigated in Golubev & Khasminskii (1999) and Golubev (2004).

Then, the functional Fourier coefficients $g_m(\cdot)$ are of the form

$$|g_m(u)| = Ku^{|m|} = K\exp(-|m|\ln(1/u)) \quad u \in [0, r_0], \tag{4.6}$$

so that $u_* = 0$, $u^* = r_0$, $|g_m(u_*)| = 0$ and $|g_m(u^*)| = K\exp(-|m|\ln(1/r_0))$ (see Example 2 in Pensky & Sapatinas (2009)). Hence, the conditions of Theorem 3 do not hold, and we cannot be certain that the convergence rates in the continuous and the discrete models coincide for any sampling scheme. Actually, it is easy to see that if sampling is carried out entirely at the single point $u_* = 0$, then $\tau_1^d(m, u_*, 1) = 0$ and we cannot recover the boundary condition $f(\cdot)$.

Example 3 Estimation of the speed of a wave on a finite interval. Let $h(t,x)$ be a solution of the wave equation

$$\frac{\partial^2 h(t,x)}{\partial t^2} = \frac{\partial^2 h(t,x)}{\partial x^2}$$

with initial–boundary conditions $h(0,x) = 0$, $\partial h(t,x)/\partial t\big|_{t=0} = f(x)$ and $h(t,0) = h(t,1) = 0$.

Here, $f(\cdot)$ is a function defined on the unit interval $[0,1]$, and the goal is to recover the speed of a wave $f(\cdot)$ on the basis of observing a noisy solution $y(t,x) = h(t,x) + n^{-1/2}z(t,x)$, where $z(t,x)$ is as in Example 1, with $t \in [a,b]$, $a > 0$, $b < 1$.

Then, the functional Fourier coefficients $g_m(\cdot)$ are of the form

$$g_0(u) = 1 \quad \text{and} \quad g_m(u) = (2\pi m)^{-1}\ \sin(2\pi m u), \quad m \in \mathbb{Z}\setminus\{0\}, \quad u \in [a,b], \tag{4.7}$$

(see Example 4 in Pensky & Sapatinas (2009)). It is easy to see that in order to satisfy the condition (4.1) the points u_* and u^* should depend on m, and, hence, the convergence rates depend on the selection of M and $\underline{u}$. Hence, the convergence rates in the continuous and the discrete models may coincide for one selection of M and $\underline{u}$ and be different for another. Actually, it is easy to see that if $M = 1$ and u is an integer, then $\tau_1^d(m, u, 1) = 0$ and we cannot recover the speed of a wave $f(\cdot)$.

4.3 Possible cases

Theorem 3 in Section 4.1 provides necessary and sufficient conditions for the convergence rates in the discrete model to be independent of the choice of M and the selection of points $\underline{u}$ and, hence, to coincide with the convergence rates in the continuous model. We can divide these conditions into the following two groups.

Condition I. There exist constants $\nu_1 \in \mathbb{R}$, $\alpha_1 \geq 0$ and $\beta_1 > 0$ and a point $u^* \in [a,b]$, independent of m and n, such that

$$|g_m(u)|^2 \leq K|g_m(u^*)|^2 \asymp |m|^{-2\nu_1}\exp(-\alpha_1|m|^{\beta_1}), \quad \nu_1 > 0 \quad \text{if} \quad \alpha_1 = 0. \tag{4.8}$$

Condition I*. There exist constants $\nu_2 \in \mathbb{R}$, $\alpha_2 \geq 0$ and $\beta_2 > 0$, and a point $u_* \in [a,b]$, independent of m and n, such that

$$|g_m(u)|^2 \geq K|g_m(u_*)|^2 \asymp |m|^{-2\nu_2}\exp(-\alpha_2|m|^{\beta_2}), \quad \nu_2 > 0 \quad \text{if} \quad \alpha_2 = 0. \tag{4.9}$$

Condition II. Either $\alpha_1\alpha_2 > 0$ and $\beta_1 = \beta_2$ or $\alpha_1 = \alpha_2 = 0$ and $\nu_1 = \nu_2$.

Consider now the following three cases.

1. **The uniform case**: Conditions I, I* and II hold.

2. **The regular case**: Condition I holds but Condition II does not hold. Condition I* holds or, possibly, $|g_m(u_*)| = 0$.

3. **The irregular case**: Condition I does not hold.

It is easy to see that Examples 1, 2 and 3 of Section 4.2 correspond to the uniform case, the regular case and the irregular case, respectively.

Theorem 3 shows that in the uniform case, the convergence rates obtained in Theorems 1 and 2 in the discrete model are independent of the choice of M and the selection of points $\underline{u}$, and, hence, coincide with the convergence rates obtained in Theorems 1 and 2 in the continuous model. In the uniform case one can replace the discrete model by the continuous model, no matter what M and $\underline{u}$ are.

In the regular case, one cannot guarantee that the convergence rates between continuous and discrete models coincide. However, as we shall show below, one can still locate a point u^* which delivers the best possible convergence rates. If sampling is done entirely at this point, then the discrete model can sometimes deliver better convergence rates than the continuous model. Nevertheless, if another sampling strategy is chosen, then the convergence rates in the discrete model may be worse than in the continuous model. Note that we do not require Condition I* to hold. This is due to the fact that Condition I* refers to the "worst case scenario" when we sample at the points which leads to the highest possible variance and, consequently, to the lowest convergence rates. One can also view $|g_m(u_*)| = 0$ as an extreme case of Condition I* when $\nu_2 = \infty$ or $\alpha_2 = \alpha_1$ and $\beta_2 = \infty$. It is easy to see that if, in the discrete model, all sampling is carried out at u_*, then the convergence rates will be worse than in the case of sampling entirely at u^* or than in the continuous model. Hence, in the regular case, sampling strategy *does* matter.

In the irregular case, it is impossible to pinpoint the best sampling strategy which suits any problem; this is due to the fact that Condition I can be violated in a variety of ways. For this reason, we study a particular example of the irregular case, namely, functional deconvolution with a box-car like blurring function; this important model occurs in the problem of estimation of the speed of a wave on a finite interval (see Example 3 in Section 4.2) and, a discretized version of it, in many areas of signal and image processing which include, for instance, LIDAR (Light Detection and Ranging) remote sensing and reconstruction of blurred images (see Section 6).

5 The regular case

5.1 The best discrete rates

It is easy to see that, in the regular case, $\tau_1^d(m, u^*, 1) \geq K \tau_1^c(m)$. Hence, it follows from Theorems 1 and 2 that, if the discrete model is sampled entirely at u^* (i.e., $M = 1$ and $u_1 = u^*$), then the asymptotical minimax lower and upper bounds for the L^2-risk in the discrete model can be only lower than the respective lower and upper bounds in the continuous model.

Denote by $\hat{f}_n^c(\cdot)$ the wavelet estimator of $f(\cdot)$ defined by (2.9) based on observations from the continuous model, and let $\hat{f}_n^d(\cdot) = \hat{f}_n^d(\underline{u}, M, \cdot)$ be the corresponding wavelet estimator of $f(\cdot)$ based on observations from the discrete model evaluated at the point $\underline{u}$. Denote $\hat{f}_n^{d*}(\cdot) = \hat{f}_n^d(u^*, 1, \cdot)$.

Then, the following statements is true.

Theorem 4 *Let $\{\phi_{j_0,k}(\cdot), \psi_{j,k}(\cdot)\}$ be the periodic Meyer wavelet basis discussed in Section 2 and assume that $s > \max(0, 1/p - 1/2)$ (for the lower bounds) or $s > 1/p'$ (for the upper bounds), $1 \leq p \leq \infty$, $1 \leq q \leq \infty$ and $A > 0$. Then,*

$$R_n^c(B_{p,q}^s(A)) \geq C R_n^d(B_{p,q}^s(A), u^*, 1) \asymp R_n^d(B_{p,q}^s(A)). \tag{5.1}$$

Also, for any choice of M and $\underline{u}$, we have

$$\sup_{f \in B^s_{p,q}(A)} \mathbb{E}\|\hat{f}_n^{d*} - f\|^2 \le C \sup_{f \in B^s_{p,q}(A)} \mathbb{E}\|\hat{f}_n^{c} - f\|^2, \tag{5.2}$$

$$\sup_{f \in B^s_{p,q}(A)} \mathbb{E}\|\hat{f}_n^{d*} - f\|^2 \le C \sup_{f \in B^s_{p,q}(A)} \mathbb{E}\|\hat{f}_n^{d} - f\|^2. \tag{5.3}$$

Theorem 4 confirms that sampling entirely at the single point u^* leads to the highest possible convergence rates in the discrete model. However, it does not provide an answer to the question whether the inequalities in (5.1) and (5.2) are strict or the convergence rates are the same in the continuous and the discrete models with sampling entirely at the single point u^*. To get a better insight into the matter, let us consider a few more examples.

5.2 More examples

Example 2 (continuation). In the case of estimation of the boundary condition for the Dirichlet problem of the Laplacian on the unit circle, the functional Fourier coefficients $g_m(\cdot)$ are of the form (4.6) with $r_0 < 1$. Hence, $u^* = r_0$ and $\tau_1^d(m, u^*, 1) \asymp |g_m(u^*)|^2 \asymp \exp(-|m| \ln(1/r_0))$. On the other hand, $\tau_1^c(m) \asymp \int_0^{r_0} u^{2|m|} du = r_0^{2|m|+1}/(2|m|+1) \asymp |m|^{-1} \exp(-|m| \ln(1/r_0))$. Hence, by Theorems 1 and 2, the convergence rates in the continuous model coincide with the convergence rates in the discrete model if sampling is carried out entirely at the single point u^*.

Example 4 Let the functional Fourier coefficients $g_m(\cdot)$ satisfy

$$|g_m(u)|^2 \asymp |m|^{-2u}, \quad 0 < a \le u \le b < \infty.$$

Then, in the continuous model, $\tau_1^c(m) = \int_a^b |g_m(u)|^2 du \asymp \int_a^b \exp(-2u \ln|m|) du \asymp |m|^{-2a} (\ln|m|)^{-1}$, implying that conditions (3.6) and (3.9) hold with $\nu = a$, $\alpha = 0$ and $\lambda = 1$. In the case of the discrete model, $u^* = a$ and $\tau_1^d(m, u^*, 1) \asymp |g_m(u^*)|^2 \asymp |m|^{-2a}$ and conditions (3.6) and (3.9) hold with $\nu = u^*$, $\alpha = 0$ and $\lambda = 0$. Hence, by Theorems 1 and 2, the convergence rates in the continuous model are worse than the convergence rates in the discrete (they differ by a logarithmic factor) model when sampling is carried out entirely at the single point u^*.

Example 5 Let the functional Fourier coefficients $g_m(\cdot)$ satisfy

$$|g_m(u)|^2 \asymp \exp\left(-\alpha |m|^u\right), \quad 0 < a \le u \le b < \infty, \tag{5.4}$$

for some constant $\alpha > 0$, independent of m. Then $u^* = a$ and $\tau_1^d(m, u^*, 1) = |g_m(u^*)|^2 \asymp \exp\left(-\alpha |m|^a\right)$. On the other hand, $\tau_1^c(m) \asymp \int_a^b \exp\left(-\alpha |m|^u\right) du \asymp (\ln|m|)^{-1} \int_{|m|^a}^{|m|^b} z^{-1} \exp(-\alpha z) dz$, so that $\tau_1^c(m) \ge K\, |m|^{-b} (\ln|m|)^{-1} \int_{|m|^a}^{|m|^b} \exp(-\alpha z) dz \asymp |m|^{-b} (\ln|m|)^{-1} \exp(-\alpha |m|^a)$ and $\tau_1^c(m) \le K\, |m|^{-a} (\ln|m|)^{-1} \int_{|m|^a}^{\infty} \exp(-\alpha z) dz \asymp |m|^{-a} (\ln|m|)^{-1} \exp(-\alpha |m|^a)$.

Hence, by Theorems 1 and 2, the convergence rates in the continuous and the discrete models coincide if sampling is carried out entirely at the single point u^*.

Example 6 Let the functional Fourier coefficients $g_m(\cdot)$ satisfy

$$|g_m(u)|^2 \asymp |m|^{-2\nu} \exp(-u|m|^{\beta}), \quad 0 \le u \le b < \infty, \tag{5.5}$$

for some constants $\nu > 0$ and $\beta > 0$, independent of m. Then, $u^* = 0$ and

$$\tau_1^d(m, u^*, 1) \asymp |g_m(u^*)|^2 \asymp |m|^{-2\nu}. \tag{5.6}$$

On the other hand, it is easy to check that

$$\tau_1^c(m) \asymp |m|^{-2\nu} \int_0^b \exp(-u|m|^{\beta}) du \asymp |m|^{-(2\nu+\beta)}. \tag{5.7}$$

Hence, by Theorems 1 and 2, the convergence rates in the continuous model are worse than in the discrete model when sampling is carried out entirely at the single point $u^* = 0$.

5.3 Conditions for convergence rates equivalency and non-equivalency between continuous and discrete models

We shall say that the convergence rates in the continuous and the discrete models "almost coincide" if the convergence rates coincide up to, at most, a logarithmic factor when the convergence rates are polynomial ($\alpha(u) \equiv 0$) or up to, at most, a constant when the convergence rates are logarithmic ($\alpha(u) > 0$). We choose this distinction between the cases of polynomial and logarithmic convergence rates since in the polynomial case the upper bounds for the risks of the adaptive estimator may differ from the corresponding lower bounds for the risk by a logarithmic factor.

Hence, a question naturally arises: under which conditions on the choice of M and the selection of sampling points $\underline{u}$ do the convergence rates in the discrete and the continuous models almost coincide and under which conditions this does not happen. In order to answer this question, first we have to derive upper and lower bounds for the L^2-risk in the continuous model.

In what follows we assume that the functional Fourier coefficients $g_m(\cdot)$ satisfy the assumption

$$|g_m(u)|^2 \asymp |m|^{-2\nu(u)} \exp\big(-\alpha(u)|m|^{\beta(u)}\big), \quad u \in U, \tag{5.8}$$

for some continuous functions $\nu(\cdot)$, $\alpha(\cdot)$ and $\beta(\cdot)$ defined on $u \in U$, such that either $\alpha(u) = 0$ and $\nu(u) > 0$ or $\alpha(u) > 0$ and $\beta(u) > 0$, for all $u \in U$. Denote

$$\vartheta = \begin{cases} \dfrac{2s}{k(2s + 2\nu(u^*) + 1)}, & \text{if} \quad \nu(u^*)(2-p) < ps^*, \\ \dfrac{2s^*}{k(2s^* + 2\nu(u^*))}, & \text{if} \quad \nu(u^*)(2-p) \geq ps^*. \end{cases} \tag{5.9}$$

Then, the following statement is valid.

Lemma 3 *Let $\{\phi_{j_0,k}(\cdot), \psi_{j,k}(\cdot)\}$ be the periodic Meyer wavelet basis discussed in Section 2 and assume that $s > \max(0, 1/p - 1/2)$ (for the lower bounds) or $s > 1/p'$ (for the upper bounds), $1 \leq p \leq \infty$, $1 \leq q \leq \infty$ and $A > 0$. Let also the functional Fourier coefficients $g_m(\cdot)$ satisfy assumption (5.8). Denote*

$$u^* = \begin{cases} \arg\min_{u \in U} \nu(u), & \text{if} \quad \alpha(u) \equiv 0, \\ \arg\min_{u \in U} \beta(u), & \text{if} \quad \alpha(u) > 0, \ \beta(u) \neq \text{const}. \end{cases}$$

Assume further that, in the neighborhood of point $u = u^*$, the function $\beta(\cdot)$ is continuously differentiable (if $\alpha(u) > 0$, $u \in U$) or the function $\nu(\cdot)$ is k-times continuously differentiable (if $\alpha(u) = 0$, $u \in U$), where $k \geq 1$ is such that

$$\nu^{(s)}(u^*) = 0, \ \ s = 1, \ldots, k-1, \ \ \nu^{(k)}(u^*) \neq 0, \tag{5.10}$$

with $\nu^{(s)}(\cdot)$ denoting the s-th derivative of the function $\nu(\cdot)$. Then, the asymptotical minimax lower and upper bounds for the L^2-risk in the continuous model are as follows:

$$R_n^c(B_{p,q}^s(A)) \geq \begin{cases} Cn^{-\frac{2s}{2s+2\nu(u^*)+1}}(\ln n)^{\vartheta}, & \text{if} \quad \alpha(u)=0,\ \nu(u^*)(2-p) < ps^*, \\ C\left(\frac{\ln n}{n}\right)^{\frac{2s^*}{2s^*+2\nu(u^*)}}(\ln n)^{\vartheta}, & \text{if} \quad \alpha(u)=0,\ \nu(u^*)(2-p) \geq ps^*, \\ C(\ln n)^{-\frac{2s^*}{\beta(u^*)}}, & \text{if} \quad \alpha(u)>0, \end{cases} \tag{5.11}$$

$$\sup_{f\in B_{p,q}^s(A)} \mathbb{E}\|\hat{f}_n^c - f\|^2 \leq \begin{cases} Cn^{-\frac{2s}{2s+2\nu(u^*)+1}}(\ln n)^{\rho+\vartheta}, & \text{if} \quad \alpha(u)=0,\ \nu(u^*)(2-p) < ps^*, \\ C\left(\frac{\ln n}{n}\right)^{\frac{2s^*}{2s^*+2\nu(u^*)}}(\ln n)^{\rho+\vartheta}, & \text{if} \quad \alpha(u)=0,\ \nu(u^*)(2-p) \geq ps^*, \\ C(\ln n)^{-\frac{2s^*}{\beta(u^*)}}, & \text{if} \quad \alpha(u)>0, \end{cases} \tag{5.12}$$

Here ρ is given by (3.14) with $\nu=\nu(u^*)$, and ϑ is given by (5.9). If $\nu(\cdot)$ is a constant function, then $k=\infty$ in (5.10) and $\vartheta=0$.

Remark 4 In Lemma 3, we do not consider the case when $\beta(u)$ is constant since this situation belongs to the uniform case and the convergence rates in the continuous and the discrete models coincide for any sampling scheme due to Theorem 3. Note also that the value of u^* in Lemma 3 is always independent of m and easy to find.

The utility of Lemma 3 is that it allows one to formulate conditions such that the convergence rates in the continuous model almost coincide with the convergence rates in the discrete model for any particular choice of a sampling scheme.

Theorem 5 *Let assumptions (5.8) and (5.10) hold.*

(i) If $\alpha(u)\equiv 0$, then the convergence rates in the continuous and the discrete models coincide up to at most a logarithmic factor if $M=M_n$ and $\underline{u}$ are such that

$$\tau_1^d(m,\underline{u},M_n) \geq K\varepsilon_n|m|^{-2\nu(u^*)}(\ln|m|)^{-\lambda_1} \tag{5.13}$$

for some constant $\lambda_1\in\mathbb{R}$, independent of m and n, and for some sequence $\varepsilon_n>0$, independent of m, satisfying

$$\lim_{n\to\infty}\varepsilon_n(\ln n)^{\lambda_2} > 0 \tag{5.14}$$

for some constant $\lambda_2\geq 0$. If, moreover, ε_n, $M=M_n$ and $\underline{u}$ are such that opposite inequalities hold, i.e.,

$$\tau_1^d(m,\underline{u},M_n) \leq C\varepsilon_n|m|^{-2\nu(u^*)}(\ln|m|)^{-\lambda_1} \quad \text{and} \quad \lim_{n\to\infty}\varepsilon_n(\ln n)^{\lambda_2} < \infty, \tag{5.15}$$

for the same constants λ_1 and λ_2 as in formulae (5.13) and (5.14), and if k in (5.10) is such that $k(\lambda_1+\lambda_2)=1$, then the convergence rates in the continuous and discrete models coincide up to constant.

(ii) If $\alpha(u)>0$, then the convergence rates in the continuous and discrete models coincide up to constant if $M=M_n$ and $\underline{u}$ are such that

$$\tau_1^d(m,\underline{u},M_n) \geq K\varepsilon_n|m|^{-2\nu}\exp\left(-\alpha|m|^{\beta(u^*)}\right)(\ln|m|)^{-\lambda_1} \tag{5.16}$$

for some constants $\nu\in\mathbb{R}$, $\lambda_1\in\mathbb{R}$ and $\alpha>0$, independent of m and n, and for some sequence $\varepsilon_n>0$, independent of m, satisfying condition (3.10).

Theorem 5 provides sufficient conditions for a sampling scheme in the discrete model to lead to the convergence rates which are optimal or near-optimal. It follows from conditions (5.8) and (5.10) and Theorems 1 and 2 that, if the discrete model is sampled entirely at u^*, then the convergence rates in the continuous and the discrete models almost coincide. Namely, as $n \to \infty$,

$$R_n^d(B_{p,q}^s(A)) \geq \begin{cases} C n^{-\frac{2s}{2s+2\nu(u^*)+1}}, & \text{if} \quad \alpha(u) = 0,\ \nu(u^*)(2-p) < ps^*, \\ C\left(\frac{\ln n}{n}\right)^{\frac{2s^*}{2s^*+2\nu(u^*)}}, & \text{if} \quad \alpha(u) = 0,\ \nu(u^*)(2-p) \geq ps^*, \\ C(\ln n)^{-\frac{2s^*}{\beta(u^*)}}, & \text{if} \quad \alpha(\mathrm{u}) > 0, \end{cases} \tag{5.17}$$

and

$$\sup_{f \in B_{p,q}^s(A)} \mathbb{E}\|\hat{f}_n^{d*} - f\|^2 \leq \begin{cases} C n^{-\frac{2s}{2s+2\nu(u^*)+1}}\ (\ln n)^{\varrho}, & \text{if} \quad \alpha(u) = 0,\ \nu(u^*)(2-p) < ps^*, \\ C\left(\frac{\ln n}{n}\right)^{\frac{2s^*}{2s^*+2\nu(u^*)}}\ (\ln n)^{\varrho}, & \text{if} \quad \alpha(u) = 0,\ \nu(u^*)(2-p) \geq ps^*, \\ C(\ln n)^{-\frac{2s^*}{\beta(u^*)}}, & \text{if} \quad \alpha(\mathrm{u}) > 0. \end{cases} \tag{5.18}$$

From the above, it also follows that

$$\frac{R_n^c(B_{p,q}^s(A))}{R_n^d(B_{p,q}^s(A))} \asymp \begin{cases} 1, & \text{if} \quad \alpha(u) > 0 \quad \text{and} \quad \beta(u) > 0, \quad u \in U, \\ (\ln n)^{\vartheta}, & \text{if} \quad \alpha(u) = 0, \quad u \in U, \end{cases}$$

and, hence, the convergence rates in the discrete model cannot be better than the convergence rates in the continuous model if $\alpha(u) > 0$ and cannot be better by more than a logarithmic factor if $\alpha(u) \equiv 0$.

We shall say that the convergence rates in the discrete model with sampling at M points $\underline{u}$ are "inferior" than the convergence rates in the continuous model if the convergence rates differ by more than a logarithmic factor for $\alpha(u) \equiv 0$ or by more than a constant factor if $\alpha(u) > 0$. The following statement shows when this happens.

Theorem 6 *Let assumptions (3.7), (5.8) and (5.10) hold and let* $\lim_{n\to\infty} \ln(\varepsilon_n)/\ln n = \varepsilon_0 < \infty$, *for some sequence* $\varepsilon_n > 0$, *independent of* m.

(i) Let $\alpha(u) \equiv 0$ *and let assumption (3.10) hold. If* $M = M_n$ *and* $\underline{u}$ *are such that*

$$\tau_1^d(m, \underline{u}, M_n) \leq K \varepsilon_n |m|^{-2\nu} (\ln |m|)^{-\lambda} \tag{5.19}$$

for some constants $\lambda \in \mathbb{R}$ *and* $\nu > 0$, *independent of* m *and* n, *then the convergence rates in the discrete model are inferior to the convergence rates in the continuous model if*

$$\nu > \nu(u^*) \ \textit{and}\ \varepsilon_0 < \begin{cases} 2(\nu - \nu(u^*))/(2s + 2\nu(u^*) + 1), & \text{if} \quad \nu(2-p) < ps^*, \\ 2(\nu - \nu(u^*))/(2s^* + 2\nu(u^*)), & \text{if} \quad \nu(2-p) \geq ps^*, \end{cases} \tag{5.20}$$

or,

$$\nu = \nu(u^*) \ \textit{and}\ \lim_{n\to\infty} \varepsilon_n (\ln n)^a = 0 \ \textit{for any} \quad a > 0. \tag{5.21}$$

(ii) Let $\alpha(u) > 0$ *and* $M = M_n$ *and* $\underline{u}$ *be such that*

$$\tau_1^d(m, \underline{u}, M_n) \leq K \varepsilon_n |m|^{-2\nu} \exp\left(-\alpha |m|^{\beta}\right) (\ln |m|)^{-\lambda} \tag{5.22}$$

for some constants $\nu \in \mathbb{R}$, $\lambda \in \mathbb{R}$ *and* $\alpha > 0$, *independent of* m *and* n. *Then, the convergence rates in the discrete model are inferior to the convergence rates in the continuous model if*

$$\beta > \beta(u^*) \ \textit{and}\ \varepsilon_0 \geq -1, \quad \textit{or} \quad \beta = \beta(u^*) \ \textit{and}\ \varepsilon_0 = -1. \tag{5.23}$$

Theorems 5 and 6 formulate conditions in terms of $\tau_1^d(m, \underline{u}, M_n)$. The following corollaries contain more specific results for various sampling schemes.

Corollary 1 *Let $M = M_n$ be finite. Then, the necessary and sufficient condition for the convergence rates in the continuous and the discrete models to almost coincide is that for at least one l, $l = 1, 2, \dots, M$, one has $\nu(u_l) = \nu(u^*)$ if $\alpha(u) \equiv 0$ or $\beta(u_l) = \beta(u^*)$ if $\alpha(u) > 0$.*

Corollary 2 *If $\alpha(u) \equiv 0$ and $M = M_n \le C(\ln n)^{\lambda^*}$ for some constant $\lambda^* \in [0, \infty)$, then the convergence rates in the continuous and the discrete models almost coincide if one has $\nu(u_l) = \nu(u^*)$ for at least one l, $l = 1, 2, \dots, M$.*

Corollary 3 *If $\alpha(u) > 0$ and $M = M_n \le Cn^\tau$ for some constant $\tau \in [0, 1)$, then the convergence rates in the continuous and the discrete models almost coincide if one has $\beta(u_l) = \beta(u^*)$ for at least one l, $l = 1, 2, \dots, M$.*

5.4 Pseudo-uniform sampling strategies

Theorems 5 and 6 and Corollaries 1, 2 and 3 in the Section 5.3 establish, in the case of an arbitrary sampling scheme, when the convergence rates in the continuous model almost coincide with the convergence rates in the discrete model, or when the convergence rates in the discrete model are inferior.

However, when the discrete model is replaced by the continuous model, the underlying implicit assumption is that sampling is carried out at $M = M_n$ equidistant points with $M_n \to \infty$. In particular, the interval $[a, b]$ is partitioned into M equal subintervals of the length $\Delta = (b-a)/M$ and $u_l = \Delta(l + d)$, $l = 0, 1, \dots, M-1$, where $d \in [0, 1]$ is the parameter which allows one to accommodate various sampling techniques (e.g., $d = 0$, $d = 1$ or $d = 1/2$, respectively, when sampling is carried out at the left, the right and the middle of each sub-interval).

Below, we study an extension of this sampling scheme. We avoid treating $u_1, u_2, \dots, u_M$ as a random sample since this is not the case in both mathematical physics and signal processing applications. Instead, in order to accommodate various sampling strategies, we consider a continuously differentiable function $S(x)$, $x \in [0, 1]$, such that $0 \le s_1 \le S'(x) \le s_2 < \infty$ and $S(0) = a$, $S(1) = b$. Let $d \in [0, 1]$ and let

$$u_l = S\left(\frac{l - 1 + d}{M}\right), \quad l = 1, 2, \dots, M. \tag{5.24}$$

Denote the inverse of $S(u)$ by $q(u) = S^{-1}(u)$, $u \in [a, b]$, and observe that $q(u)$ is continuously differentiable in $[a, b]$ with $0 \le 1/s_2 \le q'(u) \le 1/s_1 < \infty$. Many functions $S(\cdot)$ satisfy these conditions, e.g., $S(x) = a + (b-a)x^h$, where $0 < h < \infty$ (the case $h = 1$ corresponds to the uniform sampling).

Theorem 7 *Let assumptions (5.8) and (5.10) hold and let u_l, $l = 1, 2, \dots, M$, be defined by (5.24) where the function $S(x)$, $x \in [0, 1]$, is continuously differentiable such that $0 \le s_1 \le S'(x) \le s_2 < \infty$ and $S(0) = a$, $S(1) = b$. Then, the convergence rates in the discrete and the continuous models almost coincide if, for $M = M_n$,*

$$\alpha(u) \equiv 0 \text{ and } \lim_{n\to\infty} M_n^{-1} \ln n = \tau_1 < \infty, \quad \text{or} \quad \alpha(u) > 0 \text{ and } \lim_{n\to\infty} M_n^{-1} \ln\ln n = \tau_2 < \infty. \tag{5.25}$$

If, moreover, $|g_m(u)|^2 = K|m|^{-2\nu(u)}$ for some continuously differentiable function $\nu(u)$, $u \in U$, and also

$$\lim_{n\to\infty} M_n^{-1} (\ln n)^{1+1/k} = 0,$$

where k is defined in (5.10), then the convergence rates in the discrete and the continuous models coincide up to a constant.

Remark 5 Note that if $\alpha(u) > 0$ in (5.8) and d in (5.24) is such that $\beta(u_l) = \beta(u^*)$ for some l, $l = 1, 2, \ldots, M$, then a combination of Theorems 5 and 7 yields that the convergence rates in the discrete and the continuous models coincide for any value of $M = M_n$. Note also that, although conditions (5.25) in Theorem 7 are sufficient for the convergence rates in the discrete and the continuous models to almost coincide, examples in the next section demonstrate that these conditions are also necessary or close to being necessary: if the conditions in (5.25), or some slightly weaker conditions, are violated, then the convergence rates in the discrete model are inferior to the convergence rates in the continuous model.

5.5 Examples revisited

Example 2 (continuation). Recall that $|g_m(u)|^2 \asymp \exp(-2\ln(1/u)|m|)$, $u \in [0, r_0]$, so that $\beta = 1$ and $\alpha(u) = 2\ln(1/u)$. Hence, $u^* = r_0$ and if the discrete model is sampled entirely at the single point u^*, then the convergence rates in the continuous and the discrete models are given by formulae (5.11) and (5.12) or (5.17) and (5.18), respectively, and they coincide.

However, the convergence rates in the discrete and the continuous models coincide under much weaker conditions. In fact, if $M_n = O(n^\tau)$ for some constant $\tau \in [0, 1)$ and $u_l = r_1 > 0$ for at least one l, $l = 1, 2, \ldots, M$, then $\tau_1^d(m, \underline{u}, M) \geq K\, n^{-\tau} \exp(-2\ln(1/r_1)|m|)$ and, by Theorem 5, the convergence rates in the discrete and the continuous models coincide. On the other hand, if $u_1 = \cdots = u_{M-1} = 0$, $u_M = r_1 > 0$ and $M = M_n \asymp n/\ln n$, then $\tau_1^d(m, \underline{u}, M) \asymp n^{-1}\ln n\ \exp(-2\ln(1/r_1)|m|)$ and, by Theorem 6, the convergence rates in the discrete model are inferior to the convergence rates in the continuous model.

Now, consider the case of the pseudo-uniform sampling $u_l = S((l-1+d)/M)$, $l = 1, 2, \ldots, M$, with $d \in [0, 1]$ and a function $S(x)$, $x \in [0, 1]$, satisfying the assumptions of Section 5.4. We will show that the convergence rates in the discrete and the continuous models coincide no matter what the value of M is. To verify this, note that $\tau_1^d(m, \underline{u}, M) = M^{-1}\sum_{l=1}^{M} u_l^{2|m|} \leq r_0^{2|m|}$. On the other hand, it is easy to see that since $S((l-1+d)/M) = S((l-1+d)/M) - S(0) \geq s_1(l-1+d)/M$, one has $\tau_1^d(m, \underline{u}, M) \geq M^{-1}\sum_{l=M/2+1}^{M} u_l^{2|m|} \geq M^{-1}\sum_{l=M/2+1}^{M}\left(M^{-1} s_1(l-1+d)\right)^{2|m|}$. Here, $s_1 < 1$, due to $S(0) = 0$, $S(1) = r_0 < 1$ and $0 < s_1 \leq S'(x)$, and, therefore, $\tau_1^d(m, \underline{u}, M) \geq M^{-1}\sum_{l=M/2+1}^{M}(0.5\, s_1)^{2|m|} = 0.5\exp\left(-2|m|\log(2/s_1)\right)$. Since $\ln(2/s_1) > 0$, the convergence rates in the discrete and the continuous models coincide due to Theorems 1 and 2.

We conclude this example by a rather obvious observation. Reducing the sampling interval from $[0, r_0]$ to $[r_1, r_0]$, with $r_1 > 0$, yields $u_* = r_1$ and Theorem 3 immediately becomes valid. For this reason, although $|g_m(u)|$ does not satisfy condition (5.8) (since $|g_m(0)| = 0$), the convergence rates in the continuous and the discrete models coincide for the majority of "reasonable" sampling schemes. Since, with the restriction $0 < r_1 \leq u$, the problem of the estimation of the boundary condition for the Dirichlet problem of the Laplacian on the unit circle simply reduces to the uniform case, we can consider the problem as an example of an "almost uniform" case and conclude that replacing the discrete model by the continuous model is a legitimate choice.

Example 4 (continuation). Recall that $|g_m(u)|^2 \asymp |m|^{-2u}$, $u \in [a, b]$, so that $\alpha(u) = 0$, $\nu(u) = 2u$, $k = 1$ and $u^* = a$. If $M = M_n = O\left((\ln n)^{\lambda^*}\right)$ for some constant $\lambda^* \geq 0$ and $u_l = a$ for at least one l, $l = 1, 2, \ldots, M$, then, by Corollaries 1 and 2, the convergence rates in the discrete and the continuous models almost coincide. On the other hand, if $u_1 = a$ but $u_l \geq a + d$, $d > 0$,

for $l = 2, 3, \ldots, M$, and $M = M_n$ is such that $\lim_{n\to\infty} M_n (\ln n)^{-\lambda^*} = \infty$ for any constant $\lambda^* > 0$, then the convergence rates in the discrete model are inferior to those in the continuous model.

To verify this, note that under the assumptions above $\tau_1^d(m, \underline{u}, M) \leq K \left(M_n^{-1}|m|^{-2a} + |m|^{-2(a+d)}\right) \leq K \max\left(M_n^{-1}|m|^{-2a}, |m|^{-2(a+d)}\right)$. Now, apply Theorem 6, first with $\varepsilon_n = M_n^{-1}$ and $\nu = \nu(u^*)$ and then with $\varepsilon_n = 1$ and $\nu = \nu(u^*) + d$.

Now, consider the case of the pseudo-uniform sampling $u_l = S((l-1+d)/M)$, $l = 1, 2, \ldots, M$, with $d \in [0, 1]$ and a function $S(x)$, $x \in [0, 1]$, satisfying the assumptions of Section 5.4. By Theorem 7, the convergence rates in the discrete and the continuous models coincide up to, at most, a logarithmic factor if $M = M_n$ is such that $\lim_{n\to\infty} M_n^{-1} \ln n < \infty$. If, moreover, $|g_m(u)|^2 = K|m|^{-2u}$ and $M = M_n$ is such that $\lim_{n\to\infty} M_n^{-1}\ (\ln n)^2 = 0$, then the convergence rates coincide up to, at most, a constant. In other words, in each case, the convergence rates in the discrete and the continuous models almost coincide.

Let us show that the opposite is also true: if $d > 0$ and $M = M_n$ is such that

$$\lim_{n\to\infty} M_n^{-1} (\ln\ln n)^{-1} \ln n = \infty, \tag{5.26}$$

then the convergence rates in the discrete model are inferior to those in the continuous model. For this purpose, note that $u_1 - a = S(d/M) - S(0) \geq s_1 d/M = d_1/M$, so that $\tau_1^d(m, \underline{u}, M) \leq K\,|m|^{-2(a+d_1/M)} = K\,|m|^{-2a}\ \exp(-2d_1 \ln|m|/M)$. Since $\ln|m| \asymp \ln n$ in this case, we have $\tau_1^d(m, \underline{u}, M) \leq K\,\varepsilon_n\,|m|^{-2a}$ with $\varepsilon_n = \exp(-2d_1 \ln|m|/M)$. Now, the fact that the convergence rates in the discrete model are inferior to those in the continuous model follows from Theorem 6 and the observation that condition (5.26) implies condition (5.21). The latter shows that the sufficient conditions of Theorem 7 are very close to being also necessary conditions in this case.

Example 5 (continuation). Recall that $|g_m(u)|^2 \asymp \exp(-\alpha|m|^u)$, $0 < a \leq u \leq b < \infty$, so that $\alpha(u) = \alpha > 0$ and $u^* = a$. Note that, by Corollary 3, if $M = M_n$ is such that $M_n \leq Cn^\tau$ for some constant $\tau \in [0, 1)$ and $u_l = a$ for at least one l, $l = 1, 2, \ldots, M$, then the convergence rates in the discrete and the continuous models almost coincide. However, if $u_1 = a$ and $u_l \geq a + d$ for $l = 2, 3, \ldots, M$, and $M = M_n$ is such that $M_n \asymp n/\ln n$, then the convergence rates in the discrete model are inferior to those in the continuous model. To show this, note that $\tau_1^d(m, \underline{u}, M) \leq K\ \left[\exp\left(-\alpha|m|^{a+d}\right) + n^{-1}\ln n \exp\left(-\alpha|m|^a\right)\right] \asymp n^{-1}\ln n \exp\left(-\alpha|m|^a\right)$ since, in this case, $|m| \asymp (\ln n)^{1/a}$ and, thus, $\exp\left(-\alpha|m|^{a+d}\right) = o\left(n^{-1}\ln n \exp\left(-\alpha|m|^a\right)\right)$ as $n \to \infty$. Hence, application of Theorem 6 with $\varepsilon_n = \ln n/n$ yields that the convergence rates in the discrete model are inferior to the convergence rates in the continuous model.

Now, consider the case of pseudo-uniform sampling. By Theorem 7, the convergence rates in the discrete and the continuous models coincide if $M = M_n$ is such that $\lim_{n\to\infty} M_n^{-1}\ \ln\ln n < \infty$. Moreover, by Remark 5, the convergence rates in the discrete and the continuous models coincide whenever $d = 0$ in formula (5.24), no matter what the value of M is.

Let us show that, if $d > 0$, then the second condition in (5.25) is necessary in order the convergence rates in the discrete and the continuous models to coincide up to at most a constant. For this purpose, we assume that $M = M_n$ is such that $\lim_{n\to\infty} M_n^{-1}\ \ln\ln n = \infty$ and prove that the convergence rates in the discrete model are inferior to the rates in the continuous model. For this purpose, observe that $u_l \geq a + d/M$ for every l, $l = 1, 2, \ldots, M$, so that $\tau_1^d(m, \underline{u}, M) \leq K\ \exp\left(-\alpha|m|^a e^{\frac{d\ln|m|}{M}}\right)$. Now, recalling that, in this case, $\ln|m| \asymp \ln\ln n$ and $\ln n^* \asymp \ln n$, and repeating the proof of Theorem 1 with $\varepsilon_n = 1$, we obtain that, for every n, in both the sparse and the dense cases as $n \to \infty$,

$$R_n(B_{p,q}^s(A), \underline{u}, M_n) \geq C\,(\ln n)^{-\frac{2s^*}{a+d/M_n}}.$$

Hence, the convergence rates in the discrete case are inferior to those in the continuous model whenever

$$\lim_{n\to\infty} (\ln n)^{-\frac{2s^*}{a+d/M_n}+\frac{2s^*}{a}} = \lim_{n\to\infty} \exp\left(\frac{2s^* d}{a(aM_n+d)} \ln\ln n\right) = \infty,$$

which is true if $\lim_{n\to\infty} M_n^{-1} \ln\ln n = \infty$ and $d > 0$.

Example 6 (continuation). Recall that $|g_m(u)|^2 \asymp |m|^{-2\nu}\exp(-u|m|^\beta)$, $u \in [0,b]$, and that conditions of Lemma 3 do not hold since $\alpha(u) = u \geq 0$ and $\alpha(0) = 0$. We show that, in this example, the convergence rates in the discrete and the continuous models do not coincide. Recall that $u^* = 0$ and, due to formulae (5.6) and (5.7), Theorem 1 implies that, as $n \to \infty$,

$$R_n^c(B_{p,q}^s(A)) \geq \begin{cases} Cn^{-\frac{2s}{2s+2\nu+\beta+1}}, & \text{if} \quad \nu(2-p) < ps^*, \\ C\left(\frac{\ln n}{n}\right)^{\frac{2s^*}{2s^*+2\nu+\beta}}, & \text{if} \quad \nu(2-p) \geq ps^*, \end{cases} \tag{5.27}$$

and

$$R_n^d(B_{p,q}^s(A) \asymp R_n^d(B_{p,q}^s(A), u^*, 1) \geq \begin{cases} Cn^{-\frac{2s}{2s+2\nu+1}}, & \text{if} \quad \nu(2-p) < ps^*, \\ C\left(\frac{\ln n}{n}\right)^{\frac{2s^*}{2s^*+2\nu}}, & \text{if} \quad \nu(2-p) \geq ps^*, \end{cases} \tag{5.28}$$

i.e., the convergence rates, in both discrete and continuous models, are polynomial. However, if one samples the model at $u_l \geq d$, $l = 1, 2, \ldots, M$, then $\tau_1^d(m, \underline{u}, M) \leq C\,|m|^{-2\nu}\exp(-d|m|^\beta)$ and the convergence rates in the discrete model are logarithmic, i.e., as $n \to \infty$,

$$R_n^d(B_{p,q}^s(A), \underline{u}, M) \geq C(\ln n)^{-\frac{2s^*}{\beta}}. \tag{5.29}$$

Now, consider the pseudo-uniform sampling strategy $u_l = S((l-1+d)/M)$, $l = 1, 2, \ldots, M$, with a continuous differentiable function $S(x)$, $x \in [0,1]$, such that $S(0) = 0$, $S(1) = b$ and $0 \leq s_1 \leq S'(x) \leq s_2 < \infty$. Since $s_1((l-1+d)/M) \leq S((l-1+d)/M) \leq s_2((l-1+d)/M)$, $l = 1, 2, \ldots, M$, one obtains, by direct calculations, that

$$\frac{K\,|m|^{-2\nu}e^{-s_2 d|m|^\beta/M}}{M\,\left(1-e^{-s_2|m|^\beta/M}\right)} \leq \tau_1^d(m, \underline{u}, M) \leq \frac{K\,|m|^{-2\nu}e^{-s_1 d|m|^\beta/M}}{M\,\left(1-e^{-s_1|m|^\beta/M}\right)}, \tag{5.30}$$

Therefore, for $M = M_n$, the convergence rates in the discrete model depend on the value of d and the asymptotic behavior of $|m|^\beta/M_n$. Let us now show that by choosing different values of d and M_n, one can obtain each of the three convergence rates (5.27)–(5.29).

If M_n is large (e.g., $M_n \geq Cn^{1/(2\nu+\beta+1)}$), so that $|m|^\beta/M_n \to 0$ as $n \to \infty$, then $1 - e^{-s_i|m|^\beta/M_n} \asymp |m|^\beta/M_n$, $i = 1, 2$. Therefore, $\tau_1^d(m, \underline{u}, M_n) \asymp |m|^{-(2\nu+\beta)}$ and, hence, the convergence rates in the discrete and the continuous models coincide and are given by (5.27).

If M_n is small (e.g., $M_n = O(\ln n)$), so that $|m|^\beta/M_n \to \infty$ as $n \to \infty$, then (5.30) takes the form

$$K\,M_n^{-1}\,|m|^{-2\nu}e^{-s_2 d|m|^\beta/M_n} \leq \tau_1^d(m, \underline{u}, M_n) \leq K\,M_n^{-1}\,|m|^{-2\nu}e^{-s_1 d|m|^\beta/M_n}.$$

If $M = M_n$ is finite and $d > 0$, then, by Theorems 1 and 2, the convergence rates in the discrete model are logarithmic, they are given by the right hand side of formula (5.29) and are inferior to the convergence rates in the continuous model.

Finally, if $M = M_n$ is finite and $d = 0$, then the convergence rates in the discrete model are provided by the right hand side of formula (5.28) and are superior to those in the continuous model. For moderate values of M_n, one can obtain convergence rates in between (5.28) and (5.29).

6 Irregular case: a box-car like blurring function

Suppose that the blurring function $g(\cdot,\cdot)$ in the continuous model is of a box-car like, i.e.,

$$g(u,t) = 0.5\,\gamma(u)\,\mathbb{I}(|t|<u), \quad u \in U, \;\; t \in T, \tag{6.1}$$

where $\gamma(\cdot)$ is some positive function. In this case, functional Fourier coefficients $g_m(\cdot)$ satisfy

$$g_0(u) = 1 \quad \text{and} \quad g_m(u) = (2\pi m)^{-1}\,\gamma(u)\sin(2\pi m u), \quad m \in \mathbb{Z}\setminus\{0\}, \qquad u \in [a,b]. \tag{6.2}$$

It is easy to see that estimation of the initial speed of a wave on a finite interval (see Example 3 in Section 4.2) leads to $g_m(\cdot)$ of the form (6.2) with $\gamma(u)=1$ (see (4.7)).

Assume that

$$\gamma_1 \le \gamma(u) \le \gamma_2 \quad , u \in [a,b], \tag{6.3}$$

for some $0<\gamma_1\le\gamma_2<\infty$. (Obviously, this is true if $\gamma(\cdot)$ is a continuous function.) Under (6.3), it is easily seen that

$$\tau_1^c(m) \asymp m^{-2}, \tag{6.4}$$

implying that conditions (3.6) and (3.9) hold with $\nu=1$ and $\alpha=0$. Consequently, in this case, using the results of Theorems 1 and 2, we can obtain the corresponding asymptotical minimax lower and upper bounds for the L^2-risk.

Consider now the discrete model. Recall from Section 1 that this model can be viewed as a discretization of the continuous model or as a multichannel deconvolution problem with M channels where $n = NM$ denotes the total number of observations and, possibly, $M = M_n \to \infty$ as $n\to\infty$. Note that multichannel deconvolution with box-car kernels (i.e., $\gamma(u)=1/u$, for some *fixed* $u>0$) is the common problem in many areas of signal and image processing which include, for instance, LIDAR remote sensing and reconstruction of blurred images. LIDAR is a lazer device which emits pulses, reflections of which are gathered by a telescope aligned with the lazer, see, e.g., Park, Dho & Kong (1997) and Harsdorf & Reuter (2000). The return signal is used to determine distance and the position of the reflecting material. However, if the system response function of the LIDAR is longer than the time resolution interval, then the measured LIDAR signal is blurred and the effective accuracy of the LIDAR decreases. This loss of precision can be corrected by deconvolution. In practice, measured LIDAR signals are corrupted by additional noise which renders direct deconvolution impossible. Moreover, if $M\ge 2$ (finite) LIDAR devices are used to recover a signal, then we talk about a *multichannel* deconvolution problem, leading to the discrete model described by (1.2).

For any choice of M and selection of points $\underline{u}$, under (6.3), we easily see that

$$\tau_1^d(m;\underline{u},M) = \frac{1}{M}\sum_{l=1}^{M}\frac{\gamma^2(u_l)\sin^2(2\pi m u_l)}{4\pi^2 m^2} \asymp \frac{1}{m^2\,M}\sum_{l=1}^{M}\sin^2(2\pi m u_l). \tag{6.5}$$

It follows from (6.5) that for any choice of M and any selection of points $\underline{u}$, we have

$$\tau_1^d(m;\underline{u},M) \le K m^{-2}. \tag{6.6}$$

Hence, in this case, by Theorem 1, the asymptotical minimax lower bounds for the L^2-risk in this discrete model cannot be lower than the asymptotical minimax lower bounds for the L^2-risk obtained in the continuous model.

However, it is impossible to find a point $u^*\in[a,b]$, independent of m, such that, for any $u\in[a,b]$, one has $\sin^2(2\pi m u)\le K\sin^2(2\pi m u^*)$; in other words, in this case, Condition I does

not hold and we deal with the *irregular* case here. It turns out that in the case of a box-car type kernel, sampling at any one point is not at all the best strategy. Indeed, Johnstone and Raimondo (2004) showed that in the case of standard deconvolution ($M = 1$, $\gamma(u) = 1/u$, $u = u^* = a = b$), the degree of ill-posedness is $\nu = 3/2$. The latter means that the asymptotical minimax lower bounds for the L^2-risk is given by Theorem 1 with $\alpha = 0$ and $\nu = 3/2$. Johnstone and Raimondo (2004) also demonstrated that if $u^* = a$ is selected to be a 'Badly Approximable' (BA) irrational number, then these lower bounds can be attained over a wide range of ellipsiods using a nonlinear blockwise estimator in the sequence space domain.

The convergence rates obtained above can be improved by sampling at several different points. De Canditiis & Pensky (2006) studied the multichannel deconvolution problem with the box-car blurring function and derived that if M is *finite*, $M \geq 2$, one of the $u_1, u_2, \ldots, u_M$ is a BA irrational number, and $\underline{u}$ is a BA irrational tuple, then in formula (3.16)

$$\Delta_1(j) \leq C(M)\, j 2^{j(2+1/M)} \tag{6.7}$$

(for the definitions of the BA irrational number and the BA irrational tuple, see, e.g., Schmidt (1980, p. 42) and also Section 8). This implies that in this case, the degree of ill-posedness is at most $\nu \leq 1 + 1/(2M)$, meaning that if $M > 1$, then ν is less than 3/2 (that corresponds to the case of sampling at a single BA irrational number). Furthermore, De Canditiis & Pensky (2006) showed that the asymptotical upper bounds for the error (for the L^r-risk, $1 < r < \infty$, and for a *fixed* response function $f(\cdot)$) depend on M: the larger the value of M is the higher the asymptotical convergence rates will be. Hence, in the multichannel box-car deconvolution problem, it seems to be advantageous to take $M = M_n \to \infty$ as $n \to \infty$, and to choose $\underline{u}$ to be a BA irrational tuple. However, the theoretical results obtained De Canditiis & Pensky (2006) cannot be blindly applied to accommodate the case when $M = M_n \to \infty$ as $n \to \infty$; this generalization requires, possibly, non-trivial results in number theory (see the discussion in Section 8).

On the other hand, if conditions (6.1) and (6.3) hold and $M = M_n \to \infty$ fast enough as $n \to \infty$, then it is not needed to employ BA irrational tuples, as we reveal below. If $M = M_n \to \infty$ fast enough as $n \to \infty$, then deconvolution with a box-car like blurring function in the discrete model can provide estimators with the same convergence rates as in the continuous model. The following statement shows that, if $M = M_n \to \infty$ fast enough as $n \to \infty$, then an appropriate selection of points $\underline{u}$, can secure asymptotic relation similar to (6.4), thus, ensuring equal convergence rates in both the discrete and the continuous models.

Lemma 4 *Consider $g(\cdot,\cdot)$ to be of the form (6.1) with $\gamma(\cdot)$ satisfying (6.3), and let $0 < a < b < \infty$. Let $m \in A_j$, where $|A_j| = c2^j$, for some $c > 0$, with $(\ln n)^\delta \leq 2^j \leq n^{1/3}$, $j \geq j_0$, for some $\delta > 0$ and $j_0 \geq 0$. Take $u_l = a + (b-a)l/M$, $l = 1, 2, \ldots, M$. If $M \geq M_{0n} = (32\pi/3)(b-a)n^{1/3}$, then, for n and $|m|$ large enough,*

$$\tau_1^d(m; \underline{u}, M) \geq K m^{-2}. \tag{6.8}$$

Note that Lemma 4 can be applied if $M = M_n \geq c_0 n^{1/3}$ for some constant $c_0 > 0$, independent of n. Let $\Delta = \min(3c_0/(32\pi), b-a)$. Set $M = M_n$, $u_l = a + l\Delta/M$ and observe that $u_l \in [a,b]$ for $l = 1, 2, \ldots, M$. Then, the following statements is valid.

Theorem 8 *Let $\{\phi_{j_0,k}(\cdot), \psi_{j,k}(\cdot)\}$ be the periodic Meyer wavelet basis discussed in Section 2. Consider $g(\cdot,\cdot)$ to be of the form (6.1) with $\gamma(\cdot)$ satisfying (6.3), and let $0 < a < b < \infty$. Let $R_n^o(B_{p,q}^s(A))$ to be either $R_n^c(B_{p,q}^s(A))$ or $R_n^d(B_{p,q}^s(A))$.*

(Lower bounds) Let $s > \max(0, 1/p - 1/2)$, $1 \le p \le \infty$, $1 \le q \le \infty$ and $A > 0$. Then, as $n \to \infty$,

$$R_n^o(B_{p,q}^s(A)) \ge \begin{cases} Cn^{-\frac{2s}{2s+3}}, & \text{if } \ s > 3(1/p - 1/2), \\ C\left(\frac{\ln n}{n}\right)^{\frac{s'}{s'+1}}, & \text{if } \ s \le 3(1/p - 1/2). \end{cases} \tag{6.9}$$

(Upper bounds) Let $s > 1/p'$, $1 \le p \le \infty$, $1 \le q \le \infty$ and $A > 0$. Set $\nu = 1$ and assume that $M = M_n \ge c_0 n^{1/3}$ for some constant $c_0 > 0$, independent of n. Let $\hat{f}_n^c(\cdot)$ be the wavelet estimator defined by (2.9), with j_0 and J given by (3.11), and let $\hat{f}_n^d(\underline{u}, M, \cdot)$ be the wavelet estimator defined by (2.9), evaluated at the points $u_l = a + l\Delta/M$, $l = 1, 2, \ldots, M$, where $\Delta = \min(3c_0/(32\pi), b - a)$ and j_0 and J are given by (3.11). Let also $\hat{f}_n^o(\cdot)$ be either $\hat{f}_n^c(\cdot)$ or $\hat{f}_n^d(\underline{u}, M, \cdot)$. Let $s > 1/p'$, $1 \le p \le \infty$, $1 \le q \le \infty$ and $A > 0$. Then, as $n \to \infty$,

$$\sup_{f \in B_{p,q}^s(A)} \mathbb{E}\|\hat{f}_n^o - f\|^2 \le \begin{cases} Cn^{-\frac{2s}{2s+3}}\,(\ln n)^{\varrho}, & \text{if } \ s > 3(1/p - 1/2), \\ C\left(\frac{\ln n}{n}\right)^{\frac{s'}{s'+1}}\,(\ln n)^{\varrho}, & \text{if } \ s \le 3(1/p - 1/2), \end{cases} \tag{6.10}$$

where $\varrho = 3(2/p - 1)_+/(2s + 3)$ if $s > 3(1/p - 1/2)$, $\varrho = (1 - p/q)_+$ if $s = 3(1/p - 1/2)$ and $\varrho = 0$ if $s < 3(1/p - 1/2)$.

7 A limited simulation study

Here, we present a limited simulation study in the multichannel deconvolution model with a box-car like blurring function. We assess the performance of the suggested block thresholding wavelet estimator (BT) given by (2.9), with equispaced selected points $u_l = l/M$, $l = 1, 2, \ldots, M$, and compare it to the term-by-term thresholding wavelet estimator (TT) proposed by De Canditiis & Pensky (2006), where the points u_l, $l = 1, 2, \ldots, M$, were selected such that one of the u_l's is a BA irrational number, and $u_1, u_2, \ldots, u_M$ is a BA irrational tuple (see Section 4 in De Canditiis & Pensky (2006)).

Specifically, we assume that we observe

$$y(u_l, t_i) = \int_T f(x) g(u_l, t_i - x) dx + \sigma_l\, \varepsilon_{li}, \quad u_l \in U = [0, 1], \ \ t_i = i/N, \tag{7.11}$$

where $g(u_l, t) = (2u_l)^{-1}\mathbb{I}(|t| < u_l)$, $u_l \in U = [0, 1]$, and ε_{li} are standard Gaussian random variables, independent for different l and i. For simplicity, we assume that $\sigma_l^2 = \sigma^2$ for all $l = 1, 2, \ldots, M$.

The suggested algorithm consists of the following steps:

1. For each $M = 4, 8, 16$, generate M different equispaced sequences, y_{li} $(= y(u_l, i/N))$, $l = 1, 2, \ldots, M$, $i = 1, 2, \ldots, N$, following model (7.11).

2. Generate functions $g(u_l, \cdot)$, $y(u_l, \cdot)$ $\phi_{j_0 k}(\cdot)$ and $\psi_{jk}(\cdot)$, $j = j_0, j_0+1, \ldots, J-1$, $k = 0, 1, \ldots, 2^j - 1$, at the same equispaced points, $t_i = i/N$, $i = 1, 2, \ldots, N$.

3. Apply the discrete Fourier transform (FFT) to g_l, y_l, $\phi_{j_0 k}$ and ψ_{jk}, $j = j_0, j_0 + 1, \ldots, J - 1$, $k = 0, 1, \ldots, 2^j - 1$.

4. Estimate $a_{j_0 k}$ and b_{jk} by, respectively, $\hat{a}_{j_0 k}$ and $\hat{b}_{jk}$, given by (2.7).

5. Compute $\hat{B}_{jr} = \sum_{k \in U_{jr}} \hat{b}_{jk}^2$.

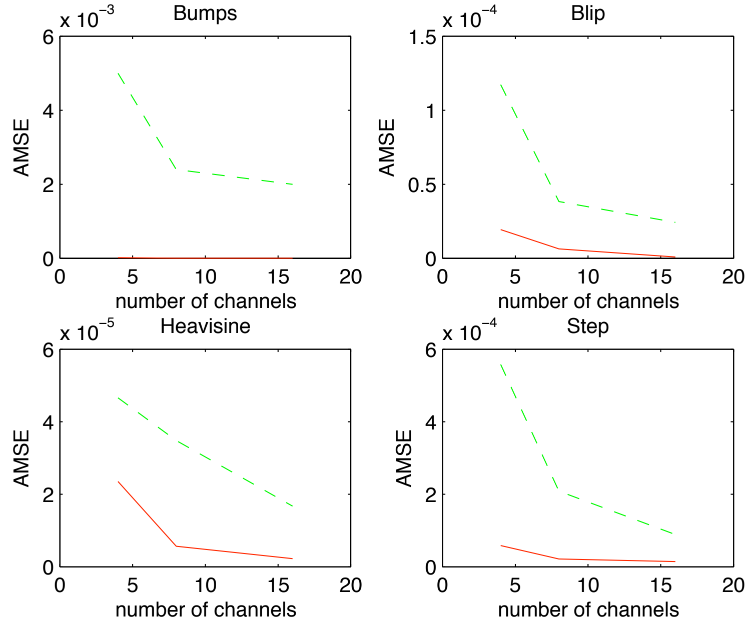

Figure 1: AMSE for the *Bumps*, *Blip*, *Heavisine* and *Step* functions sampled at a fixed number of $N = 128$ points, based on RSNR=1, as the number of channels M (and hence the sample size n) increases. Solid line: BT wavelet estimator; Dash line: TT wavelet estimator.

6. Compute the threshold $\lambda_j = \hat{\sigma}^2 d^* n^{-1} \ln n \, \Delta_1(j), \quad j \geq j_0$, where $n = NM$, $d^* = 1$,

$$\hat{\sigma} = \sqrt{\frac{1}{M(N-2)} \sum_{l=1}^{M} \sum_{i=2}^{N-1} \left(\frac{y_{l,i-1}}{\sqrt{6}} - \frac{2y_{li}}{\sqrt{6}} + \frac{y_{l,i+1}}{\sqrt{6}} \right)^2}, \quad \Delta_1(j) = \frac{1}{|C_j|} \sum_{m \in C_j} \tau_1^{-1}(m).$$

(see Remark 6 in Pensky & Sapatinas (2009) and Müller & Stadmüller (1987)).

7. Threshold the wavelet coefficients belonging to blocks with $|\hat{B}_{jr}| < \lambda_j$.

8. Apply the inverse wavelet transform to obtain $\hat{f}_n(\cdot)$ given by (2.9).

We used the test functions "Bumps", "Blip", "Heavisine" and "Step", and set $j_0 = 3$. For a fixed value of the (root) signal-to-noise ratio (RSNR=1), we generated $S = 100$ samples of size $n = NM$ from model (7.11) in order to calculate the average mean-squared error (AMSE) given by

$$S^{-1} \sum_{m=1}^{S} \sum_{i=1}^{N} (\hat{f}_n^m(t_i) - f(t_i))^2 \Bigg/ \sum_{i=1}^{N} f^2(t_i), \quad t_i = i/N.$$

In Figure 1, for a fixed number of data points $N = 2^7$, we evaluate the AMSE as the number of channels M, and hence the sample size n, increases for the four signals mentioned above. Obviously, both BT and TT wavelet estimators improve their performances, as n increases, and the BT wavelet estimator appears to have smaller AMSE than the TT wavelet estimator in all cases.

Although not reported here, we also evaluated precision of the suggested BT wavelet estimator for a wide variety of other test functions (see the list of test functions in Appendix I of

Antoniadis, Bigot & Sapatinas (2001)) and RSNR's, with very good performances. This numerical study confirms that under the multichannel deconvolution model with a box-car like blurring function, block thresholding wavelet estimators with equispaced selection of points u_l, $l = 1, 2, \ldots, M$, produce quite accurate estimates of $f(\cdot)$.

8 Concluding remarks

We considered the question of whether and when, in the functional deconvolution setting, it is legitimate to replace the real-life discrete deconvolution problem by its continuous idealization. In other words, using the asymptotical minimax framework, we studied whether the continuous model and the discrete model are equivalent for some or any sampling schemes from the viewpoint of convergence rates, over a wider range of Besov balls and for the L^2-risk. It is worth mentioning that when we talked about convergence rates we referred to the lower bounds which are attainable up to, at most, a logarithmic factor according to Theorems 1 and 2. In the cases when convergence rates in the discrete model depend on the choice of a sampling scheme, we also explored the optimal sampling strategies. The conclusions of our investigation can be summarized as follows.

If Conditions I, I* and II are satisfied then the convergence rates in the discrete model are independent of the number M and the choice of sampling points $\underline{u}$ and coincide with the convergence rates in the continuous model. In this case, which we call *uniform*, it is legitimate to replace discrete model (with any selection of sampling points) by continuous model.

If Condition II does not hold, then there exist at least two different sampling schemes in discrete model which deliver two different sets of convergence rates, and at least one of these sampling schemes leads to the convergence rates different from the continuous model. However, if Condition I holds, one can point out the sampling scheme which delivers the fastest convergence rates, namely, sampling entirely at "the best possible" point u^*. We refer to this case as *regular* and explore when, under an arbitrary sampling scheme, convergence rates in the discrete model coincide or do not coincide with the convergence rates in the continuous model. The case of sampling at u^* is studied as a particular case.

In addition, we consider convergence rates in the discrete model under uniform or pseudo-uniform sampling strategies. Indeed, when a discrete model is replaced by its continuous counterpart, it is implicitly assumed that sampling is carried out at M equidistant points in the interval $[a, b]$. We formulate conditions when this replacement is legitimate and bring examples when the uniform, or a more general pseudo-uniform, sampling may lead to convergence rates which differ from the convergence rates in the continuous model and are lower than when sampling is carried out entirely at the "best possible" point u^*. Hence, even in the regular case, one should be extremely careful when replacing a discrete model by its continuous counterpart.

Finally, we study the case when Condition I is violated. We referred to this case as *irregular*. In this case, the convergence rates in the discrete model depend on a sampling strategy and, in addition, one cannot design a sampling scheme which delivers the highest convergence rates. Since Condition I can be violated in a variety of ways, in the irregular case a general study is very complex. For this reason, we study a particular example of the irregular case, namely, functional deconvolution with a box-car like blurring function. This important model occurs, in the problem of estimation of the speed of a wave on a finite interval (Example 3) as well as, a discrete version of it, in signal and image processing (see Section 6). In the case of a box-car like kernel, sampling at any one point is, by far, not the best possible choice and delivers lower convergence rates than the continuous model. The best choice for this model is uniform sampling with a large value of $M = M_n$. Indeed, if $M = M_n \geq c_0 n^{1/3}$ for some constant $c_0 > 0$, independent of n, and the selection points $u_1, u_2, \ldots, u_M$, are selected to be equispaced, then, according to Theorem 8, the

convergence rates in the discrete model with a box-car like blurring function coincide with the convergence rates in the continuous model and cannot be improved.

The assumption that $M = M_n$ grows at least at a rate of $n^{1/3}$ is very natural in the inverse mathematical physics problems: in fact, if one samples uniformly in the rectangle $[0,1] \times [a,b]$, then $M_n \asymp \sqrt{n}$. However, this assumption is hardly natural in a signal processing setting where M corresponds to a number of physical devices, so even if $M = M_n \to \infty$ as $n \to \infty$, it grows at a very slow rate. For this reason, the question remains: if $M = M_n \to \infty$ at a rate slower than $O(n^{1/3})$ (e.g., $M = M_n = c_3 n^{\upsilon}$, where $0 < \upsilon < 1/3$, or $M = M_n = c_4(\ln n)^{\gamma}$, where $\gamma > 0$, for some constants $c_3 > 0$ and $c_4 > 0$, independent of n), can one select points $u_l \in [a,b]$, $l = 1, 2, \ldots, M$, such that the convergence rates in the discrete model coincide with the corresponding convergence rates obtained in the continuous model? And, if for some such $M = M_n$ the convergence rates in the discrete and the continuous models are not the same, what are the best convergence rates that can be attained and the best selection of points $u_1, u_2, \ldots, u_M$?

The solution of this question, possibly, rests on very nontrivial results in number theory. Recall that De Canditiis & Pensky (2006) showed that, if M is finite, $M \geq 2$, one of the u_l's is a BA irrational number, and $u_1, u_2, \ldots, u_M$ is a BA irrational tuple, then (6.7) is valid. The constant $C(M)$ in (6.7) depends on the value of M and the choice of the BA irrational tuple. Let us now elaborate more on this. Note that the numbers $a_1, a_2, \ldots, a_M$ is a BA irrational tuple (see, e.g., Schmidt (1980, p. 42)), if, for any integers $p_1, p_2, \ldots, p_M$ and q, there exists constant B_M such that

$$\max\left(|a_1 q - p_1|, |a_2 q - p_2|, \ldots, |a_M q - p_1|\right) \geq B_M q^{-1/M},$$

where B_M is a positive constant that depends only on M. Schmidt (1980, p. 43) showed that, for a finite value of M, a BA irrational tuple always exist, and proposed an algorithm for constructing it. It is easy to note that $B_M \to 0$ as $M \to \infty$. The value of B_M affects the value of $C(M)$ in (6.7) and, therefore, the convergence rates in the discrete model.

Unfortunately, we are not aware of any results in number theory on how B_M depends on M, and we suspect that relevant results may not have yet been derived. However, a partial answer to the above question, showing that $B_M \geq C_0 \exp(-3M \ln M)$, for some $C_0 > 0$, independent of M, q and $p_1, p_2, \ldots, p_M$, and the construction of minimax upper bounds for the L^2-risk over a wide range of Besov balls, covering the case $M = M_n = o((\ln n)^u)$, where $u \geq 1/2$, have been recently obtained in Pensky & Sapatinas (2009a).

9 Appendix: Proofs

Recall that the symbol C is used for a generic positive constant, independent of n, while the symbol K is used for a generic positive constant, independent of m, n, M and $u_1, u_2, \ldots, u_M$, which either of them may take different values at different places.

Proof of Theorem 1. The proof of the lower bounds falls into two parts. First, we consider the lower bounds obtained when the worst functions f (i.e., the hardest functions to estimate) are represented by only one term in a wavelet expansion (sparse case), and then when the worst functions f are uniformly spread over the unit interval T (dense case).

In the continuous model, one can always choose $\varepsilon_n = 1$, so the only difference with Pensky & Sapatinas (2009) is an extra logarithmic factor. Since the differences for the discrete model are much more significant, we only consider below the proof for the discrete model.

Sparse case. Let the functions f_{jk} be of the form $f_{jk} = \gamma_j \psi_{jk}$ and let $f_0 \equiv 0$. Note that by (3.1), in order $f_{jk} \in B^s_{p,q}(A)$, we need $\gamma_j \leq A 2^{-js'}$. Set $\gamma_j = c 2^{-js'}$, where c is a positive constant

such that $c < A$, and apply the following classical lemma on lower bounds:

Lemma 5 (Härdle *et.al.* (1998), Lemma 10.1). *Let V be a functional space, and let $d(\cdot,\cdot)$ be a distance on V. For $f, g \in V$, denote by $\Lambda_n(f,g)$ the likelihood ratio $\Lambda_n(f,g) = d\mathbb{P}_{X_n^{(f)}}/d\mathbb{P}_{X_n^{(g)}}$, where $d\mathbb{P}_{X_n^{(h)}}$ is the probability distribution of the process X_n when h is true. Let V contains the functions $f_0, f_1, \ldots, f_\aleph$ such that (a) $d(f_k, f_{k'}) \geq \delta > 0$ for $k = 0, 1, \ldots, \aleph$, $k \neq k'$; (b) $\aleph \geq \exp(\lambda_n)$ for some $\lambda_n > 0$; (c) $\ln \Lambda_n(f_0, f_k) = u_{nk} - v_{nk}$, where v_{nk} are constants and u_{nk} is a random variable such that there exists $\pi_0 > 0$ with $\mathbb{P}_{f_k}(u_{nk} > 0) \geq \pi_0$; (d) $\sup_k v_{nk} \leq \lambda_n$.*

Then, $\sup_{f\in V} \mathbb{P}_{X_n^{(f)}}\big(d(\tilde{f}, f) \geq \delta/2\big) \geq \pi_0/2$ for any arbitrary estimator $\tilde{f}$.

Let now $V = \left\{f_{jk} : 0 \leq k \leq 2^j - 1\right\}$ so that $\aleph = 2^j$. Choose $d(f,g) = \|f - g\|$, where $\|\cdot\|$ is the L^2-norm on the unit interval T. Then, $d(f_{jk}, f_{jk'}) = \gamma_j = \delta$. Let $v_{nk} = \lambda_n = j \ln 2$ and $u_{nk} = \ln \Lambda_n(f_0, f_{jk}) + j\ln 2$. Now, to apply Lemma 5, we need to show that for some $\pi_0 > 0$, uniformly for all f_{jk}, we have $\mathbb{P}_{f_{jk}}(u_{nk} > 0) = \mathbb{P}_{f_{jk}}\big(\ln \Lambda_n(f_0, f_{jk}) > -j\ln 2\big) \geq \pi_0 > 0$.

Note that in the case of discrete model

$$-\ln \Lambda_n(f_0, f_{jk}) = 0.5 \sum_{i=1}^{N}\sum_{l=1}^{M} \left\{[y(u_l,t_i) - \gamma_j(\psi_{jk} * g)(u_l,t_i)]^2 - y^2(u_l,t_i)\right\} = v_{jk} - u_{jk},$$

where

$$u_{jk} = \gamma_j \sum_{i=1}^{N}\sum_{l=1}^{M} (\psi_{jk} * g)(u_l,t_i)\,\varepsilon_{li}, \quad v_{jk} = 0.5\,\gamma_j^2 \sum_{i=1}^{N}\sum_{l=1}^{M} [(\psi_{jk} * g)(u_l,t_i)]^2.$$

Observe that, due to $\mathbb{P}(\varepsilon_{li} > 0) = \mathbb{P}(\varepsilon_{li} < 0) = 0.5$, we have $\mathbb{P}(u_{jk} > 0) = 0.5$. By properties of the discrete Fourier transform and taking into account that in the case of Meyer wavelets, $|\psi_{mjk}| \leq 2^{-j/2}$ (see, e.g., Johnstone *et.al.* (2004), p. 565), we derive that

$$v_{jk} \leq \frac{\gamma_j^2}{4\pi} \sum_{i=1}^{N}\sum_{l=1}^{M}\sum_{m\in C_j} |\psi_{mjk}|^2 |g_m(u_l)|^2 \leq \frac{NM\gamma_j^2}{4\pi 2^j} \sum_{m\in C_j} M^{-1} \sum_{l=1}^{M} |g_m(u_l)|^2 \equiv B_n,$$

where $B_n = (4\pi)^{-1}\, n 2^{-j}\gamma_j^2 \, \sum_{m\in C_j} \tau_1^d(m, \underline{u}, M)$.

Let $j = j_n$ be such that $B_n \leq 0.5\, j\ln 2$. Then, by applying Lemma 5 and Chebyshev's inequality, we obtain

$$\inf_{\tilde{f}_n} \sup_{f\in B_{p,q}^s(A)} \mathbb{E}\|\tilde{f}_n - f\|^2 \geq \inf_{\hat{f}_n} \sup_{f\in V} \frac{1}{4}\gamma_j^2 \mathbb{P}\big(\|\tilde{f}_n - f\| \geq \gamma_j/2\big) \geq 0.25\,\gamma_j^2\pi_0. \tag{9.1}$$

Thus, we just need to choose the smallest possible $j = j_n$ satisfying $B_n \leq 0.5\, j\ln 2$, evaluate $\gamma_j = c2^{-js'}$, and to plug it into (9.1). By direct calculations, we derive, under condition (3.6), that

$$\sum_{m\in C_j} \tau_1^d(m, \underline{u}, M) \leq \begin{cases} K\varepsilon_n \; 2^{-j(2\nu-1)} j^{-\lambda}, & \text{if } \alpha = 0, \\ \\ K\varepsilon_n \; 2^{-j(2\nu+\beta-1)} j^{-\lambda} \; \exp\left(-\alpha(2\pi/3)^\beta 2^{j\beta}\right), & \text{if } \alpha > 0. \end{cases} \tag{9.2}$$

Hence, if $\alpha = 0$, then $2^{j_n} = C\left(n^*(\ln n^*)^{-(\lambda+1)}\right)^{1/(2s'+2\nu)}$ and, if $\alpha > 0$, then $2^{j_n} = C(\ln n^*)^{1/\beta}$. Now, to obtain the lower bound, plug $\gamma_j = c2^{-j_n s'}$ into (9.1)

$$\inf_{\tilde{f}_n} \sup_{f\in B_{p,q}^s} \mathbb{E}\|\tilde{f}_n - f\|^2 \geq \begin{cases} C(n^*)^{\frac{2s'}{2s'+2\nu}} \; (\ln n^*)^{\frac{2s'(\lambda+1)}{2s'+2\nu}}, & \text{if } \alpha = 0, \\ \\ C(\ln n^*)^{-\frac{2s'}{\beta}}, & \text{if } \alpha > 0. \end{cases} \tag{9.3}$$

.

Dense case. Let η be the vector with components $\eta_k = \pm 1$, $k = 0, 1, \ldots, 2^j - 1$, denote by Ξ the set of all possible vectors η, and let $f_{j\eta} = \gamma_j \sum_{k=0}^{2^j-1} \eta_k \psi_{jk}$. Let also η^i be the vector with components $\eta^i_k = (-1)^{\mathbb{I}(i=k)} \eta_k$ for $i, k = 0, 1, \ldots, 2^j - 1$. Note that by (3.1), in order $f_{j\eta} \in B^s_{p,q}(A)$, we need $\gamma_j \le A 2^{-j(s+1/2)}$. Set $\gamma_j = c_\star 2^{-j(s+1/2)}$, where $c_\star$ is a positive constant such that $c_\star < A$, and apply the following lemma on lower bounds:

Lemma 6 (Willer (2005), Lemma 2). *Let $\Lambda_n(f,g)$ be defined as in Lemma 5, and let η and $f_{j\eta}$ be as described above. Suppose that, for some positive constants λ and π_0, we have $\mathbb{P}_{f_{j\eta}}(-\ln \Lambda_n(f_{j\eta^i}, f_{j\eta}) \le \lambda) \ge \pi_0$, uniformly for all $f_{j\eta}$ and all $i = 0, \ldots, 2^j - 1$. Then, for any arbitrary estimator $\tilde f$ and for some constant $L > 0$, one has* $\max_{\eta \in \Xi} \mathbb{E}_{f_{j\eta}} \|\tilde f - f_{j\eta}\| \ge L \pi_0 e^{-\lambda} 2^{j/2} \gamma_j$.

Since, by Chebychev's inequality, $\mathbb{P}_{f_{jk}}\big(\ln \Lambda_n(f_{j\eta^i}, f_{j\eta}) > -\lambda\big) \ge 1 - \mathbb{E}_{f_{jk}} \big|\ln \Lambda_n(f_{j\eta^i}, f_{jk})\big| / \lambda$, we need to show that $\mathbb{E}_{f_{j\eta}} |\ln \Lambda_n(f_{j\eta^i}, f_{j\eta})| \le \lambda_1$, for a sufficiently small constant $\lambda_1 > 0$. Observe that

$$\ln \Lambda_n(f_{j\eta^i}, f_{j\eta}) = 0.5 \gamma_j^2 \sum_{i=1}^N \sum_{l=1}^M [(g * f_{j\eta^i} - f_{j\eta})(u_l, t_i)]^2 - \gamma_j \sum_{i=1}^N \sum_{l=1}^M \varepsilon_{li} \left[(g * [f_{j\eta^i} - f_{j\eta}])(u_l, t_i)\right]$$

Then, due to $|f_{j\eta^i} - f_{j\eta})| = 2|\psi_{jk}|$, one has $\mathbb{E}_{f_{j\eta}} |\ln \Lambda_n(f_{j\eta^i}, f_{j\eta})| \le A_n + B_n$, where

$$A_n = 2\gamma_j \, \mathbb{E} \Big| \sum_{i=1}^N \sum_{l=1}^M (\psi_{jk} * g)(u_l, t_i) \varepsilon_{li} \Big|, \quad B_n = 2\gamma_j^2 \, \sum_{i=1}^N \sum_{l=1}^M (\psi_{jk} * g)^2 (u_l, t_i).$$

Since, by Jensen's inequality, $A_n \le \sqrt{2B_n}$, we only need to construct an upper bound for B_n. Note that, similarly to the sparse case, one has $B_n = O(n 2^{-j} \gamma_j^2 \, \sum_{m \in C_j} \tau_1^d(m, \underline{u}, M))$. According to Lemma 6, we choose $j = j_n$ that satisfies the condition $B_n + \sqrt{2B_n} \le \lambda_1$. Using (9.2), we derive that $2^{j_n} = C\left(n^* (\ln n^*)^{-\lambda}\right)^{1/(2s+2\nu+1)}$ if $\alpha = 0$ and $2^{j_n} = C (\ln n^*)^{1/\beta}$ if $\alpha > 0$. Then, Lemma 6 and Jensen's inequality yield

$$\inf_{\tilde f_n} \sup_{f \in B^s_{p,q}} \mathbb{E} \|\tilde f_n - f\|^2 \ge \begin{cases} C (n^*)^{-\frac{2s}{2s+2\nu+1}} \, (\ln n^*)^{-\frac{2s\lambda}{2s+2\nu+1}}, & \text{if} \quad \alpha = 0, \\ \\ C (\ln n^*)^{-\frac{2s}{\beta}}, & \text{if} \quad \alpha > 0. \end{cases} \tag{9.4}$$

Now, to complete the proof one just needs to note that $s^* = \min(s, s')$, and that

$$2s/(2s + 2\nu + 1) \le 2s^*/(2s^* + 2\nu) \quad \text{if} \quad \nu(2-p) \le p s^*, \tag{9.5}$$

with the equalities taken place simultaneously, and then to choose the highest of the lower bounds (9.3) and (9.4). This completes the proof of Theorem 1. □

Proof of Lemma 1. In what follows, we shall only construct the proof for b_{jk} since the proof for $a_{j_0 k}$ is very similar. Again, we construct the proof only for discrete model, since in the case of continuous model, one can always choose $\varepsilon_n = 1$, so the only difference with Pensky & Sapatinas (2009) is an extra logarithmic factor.

Note that, by (2.7), one has $\widehat{b}_{jk} - b_{jk} = \sum_{m \in C_j} (\widehat{f}_m - f_m) \overline{\psi_{mjk}}$, with

$$\widehat{f}_m - f_m = N^{-1/2} \, \Big(\sum_{l=1}^M \overline{g_m(u_l)} \, z_{ml} \Big) \Big/ \Big(\sum_{l=1}^M |g_m(u_l)|^2 \Big), \tag{9.6}$$

where z_{ml} are standard Gaussian random variables, independent for different m and l. Therefore,

$$\mathbb{E}|\widehat{b}_{jk} - b_{jk}|^2 = N^{-1} \sum_{m \in C_j} |\psi_{mjk}|^2 \left[\sum_{l=1}^{M} |g_m(u_l)|^2 \right]^{-1} = O\Big(n^{-1} \Delta_1(j)\Big)$$

since $|C_j| = 4\pi 2^j$ and $|\psi_{mjk}|^2 \le 2^{-j}$. If $\kappa = 2$, then

$$\begin{aligned} \mathbb{E}|\widehat{b}_{jk} - b_{jk}|^4 &= O\Bigg(\sum_{m \in C_j} \mathbb{E}|\widehat{f}_m - f_m|^4 |\psi_{mjk}|^4 \Bigg) + O\Bigg(\Bigg[\sum_{m \in C_j} \mathbb{E}|\widehat{f}_m - f_m|^2 |\psi_{mjk}|^2 \Bigg]^2 \Bigg) \\ &= O\Big(2^{-j} N^{-2} M^{-3} \Delta_2(j) + N^{-2} M^{-2} \Delta_1^2(j)\Big) = O\Big(2^{-j} n^{-2} M^{-1} \Delta_2(j) + n^{-2} \Delta_1^2(j)\Big). \end{aligned} \tag{9.7}$$

Direct calculations show that when $\alpha = 0$ one has $\Delta_2(j) = O(2^{6j\nu} j^{3\lambda} \varepsilon_n^{-3})$. Plugging expressions for $\Delta_1(j)$ and $\Delta_2(j)$ into formula (9.7) and taking into account that $2^j \le 2^{J-1} < (n^*)^{1/(2\nu+1)}$, one derives

$$\mathbb{E}|\widehat{b}_{jk} - b_{jk}|^4 = O\left(\frac{2^{6j\nu} j^{3\lambda}}{n^2 \varepsilon^3 M_n} + \frac{2^{4j\nu} j^{2\lambda}}{n^2 \varepsilon^2} \right) = O\left(n\ (n^*)^{\frac{6\nu}{2\nu+1} - 3} (\ln n)^{3\lambda} + (n^*)^{\frac{4\nu}{2\nu+1} - 2} (\ln n)^{2\lambda} \right).$$

To complete the proof, observe that in the last expression, the second term is asymptotically smaller than the first. $\square$

Proof of Lemma 2. Again we carry out the proof only for discrete case. The proof for continuous case can be obtain as a minor variation of the proof below. Consider the set of vectors $\Omega_{jr} = \Big\{ v_k,\ k \in U_{jr} :\ \sum_{k \in U_{jr}} |v_k|^2 \le 1 \Big\}$, and the centered Gaussian process defined by $Z_{jr}(v) = \sum_{k \in U_{jr}} v_k (\widehat{b}_{jk} - b_{jk})$. The proof of the lemma is based on the following inequality:

Lemma 7 (Cirelson, Ibragimov & Sudakov (1976)). *Let D be a subset of $\mathbb{R} = (-\infty, \infty)$, and let $(\xi_t)_{t \in D}$ be a centered Gaussian process. If $\mathbb{E}(\sup_{t \in D} \xi_t) \le B_1$ and $\sup_{t \in D} \mathrm{Var}(\xi_t) \le B_2$, then, for all $x > 0$, we have $\mathbb{P}\Big(\sup_{t \in D} \xi_t \ge x + B_1 \Big) \le \exp\Big(-x^2/(2B_2) \Big)$.*

To apply Lemma 7, we need to find B_1 and B_2. Note that, by Jensen's inequality, we obtain

$$\mathbb{E}\left[\sup_{v \in \Omega_{jr}} Z_{jr}(v) \right] = \mathbb{E}\left[\sum_{k \in U_{jr}} |\widehat{b}_{jk} - b_{jk}|^2 \right]^{1/2} \le \left[\sum_{k \in U_{jr}} \mathbb{E}|\widehat{b}_{jk} - b_{jk}|^2 \right]^{1/2} \le \frac{\sqrt{c_1} 2^{\nu j} j^{\lambda/2} \sqrt{\ln n}}{\sqrt{n^*}} = B_1.$$

(Here, c_1 is the same positive constant as in (3.16) with $\alpha = 0$.) Also, by (2.2) and (9.6), we have $\mathbb{E}\big[(\widehat{b}_{jk} - b_{jk})(\widehat{b}_{jk'} - b_{jk'})\big] = n^{-1} \sum_{m \in C_j} \psi_{mjk} \overline{\psi_{mjk'}} [\tau_1(m)]^{-1}$, where $\tau_1(m)$ is defined in (3.5). Hence,

$$\begin{aligned} \sup_{v \in \Omega_{jr}} \mathrm{Var}(Z_{jr}(v)) &= n^{-1} \sup_{v \in \Omega_{jr}} \sum_{k \in U_{jr}} \sum_{k' \in U_{jr}} v_k v_{k'} \sum_{m \in C_j} \psi_{mjk} \overline{\psi_{mjk'}} [\tau_1(m)]^{-1} \\ &\le c_1 (n^*)^{-1} 2^{2\nu j} j^\lambda \sum_{k \in U_{jr}} v_k^2 \le c_1 (n^*)^{-1} 2^{2\nu j} j^\lambda = B_2, \end{aligned}$$

by using $\sum_{m \in C_j} \psi_{mjk} \overline{\psi_{mjk'}} = \mathbb{I}(k = k')$ and (3.16) for $\alpha = 0$. Therefore, by applying Lemma 7 with B_1 and B_2 defined above and $x = B_1 \left((2\sqrt{c_1})^{-1} \mu \sqrt{h_2} - 1 \right)$, and noting that under condition (3.19), $\ln(n^*) \ge h_2 \ln n$, we derive

$$\mathbb{P}\Bigg(\sum_{k \in U_{jr}} |\widehat{b}_{jk} - b_{jk}|^2 \ge \frac{\mu^2\, 2^{2\nu j}\, j^\lambda\, \ln(n^*)}{4n^*} \Bigg) \le \exp\left\{ -\left(\frac{\mu \sqrt{h_2}}{2\sqrt{c_1}} - 1 \right)^2 \frac{B_1^2}{2B_2} \right\} \le n^{-3},$$

since (3.19) implies that $0.5\ [\mu\sqrt{h_2}/(2\sqrt{c_1})-1]^2 \geq 3$. This completes the proof of Lemma 2. $\square$

Proof of Theorem 2. First, note that in the case of $\alpha > 0$, we have $\mathbb{E}\|\hat{f}_n - f\|^2 = R_1 + R_2$, where

$$R_1 = \sum_{j=J}^{\infty} \sum_{k=0}^{2^j-1} b_{jk}^2, \qquad R_2 = \sum_{k=0}^{2^{j_0}-1} \mathbb{E}(\widehat{a}_{j_0 k} - a_{j_0 k})^2, \tag{9.8}$$

since $j_0 = J$. It is well-known (see, e.g., Johnstone (2002), Lemma 19.1) that if $f \in B^s_{p,q}(A)$, then for some positive constant $c^\star$, dependent on p, q, s and A only, we have

$$\sum_{k=0}^{2^j-1} b_{jk}^2 \leq c^\star 2^{-2js^*}, \tag{9.9}$$

thus, $R_1 = O\left(2^{-2Js^*}\right) = O\left((\ln n^*)^{-2s^*/\beta}\right)$. Also, using (3.16) and (3.17), we derive $R_2 = O\left(n^{-1} 2^{j_0} \Delta_1(j_0)\right) = O\left((n^*)^{-1/2} (\ln n^*)^{2\nu/\beta}\right) = o\left((\ln n^*)^{-2s^*/\beta}\right)$, thus completing the proof for $\alpha > 0$.

Now, consider the case of $\alpha = 0$. Note that by condition (3.10) one has $\ln n^* \asymp \ln n$. Due to the orthonormality of the wavelet basis, we obtain

$$\mathbb{E}\|\hat{f}_n - f\|^2 = R_1 + R_2 + R_3 + R_4, \tag{9.10}$$

where R_1 and R_2 are defined in (9.8), and

$$R_3 = \sum_{j=j_0}^{J-1} \sum_{r \in A_j} \sum_{k \in U_{jr}} \mathbb{E}\left[(\widehat{b}_{jk} - b_{jk})^2\ \mathbb{I}(\widehat{B}_{jr} \geq \mu^2\ (n^*)^{-1} 2^{2\nu j} \ln(n^*) j^\lambda)\right],$$

$$R_4 = \sum_{j=j_0}^{J-1} \sum_{r \in A_j} \sum_{k \in U_{jr}} \mathbb{E}\left[b_{jk}^2\ \mathbb{I}(\widehat{B}_{jr} < \mu^2\ (n^*)^{-1} 2^{2\nu j} \ln(n^*) j^\lambda)\right],$$

where $\widehat{B}_{jr}$ and μ are defined by (2.8), (3.13) and (3.19), respectively.

Let us now examine each term in (9.10) separately. Similarly to the case of $\alpha > 0$, we obtain $R_1 = O\left(2^{-2Js^*}\right) = O\left(n^{-2s^*/(2\nu+1)}\right)$. By direct calculations, one can check that $2s^*/(2\nu+1) > 2s/(2s+2\nu+1)$, if $\nu(2-p) < ps^*$, and $2s^*/(2\nu+1) \geq 2s^*/(2s^*+2\nu)$, if $\nu(2-p) \geq ps^*$. Hence,

$$R_1 = \begin{cases} O\left((n^*)^{-\frac{2s}{2s+2\nu+1}}\right), & \text{if} \quad \nu(2-p) < ps^*, \\ O\left((n^*)^{-\frac{2s^*}{2s^*+2\nu}}\right), & \text{if} \quad \nu(2-p) \geq ps^*. \end{cases} \tag{9.11}$$

Also, by (3.17) and (3.16), we obtain

$$R_2 = O\left((n^*)^{-1} 2^{(2\nu+1)j_0}\right) = o\left((n^*)^{-\frac{2s}{2s+2\nu+1}}\right) = o\left((n^*)^{-\frac{2s^*}{2s^*+2\nu}}\right). \tag{9.12}$$

Denote

$$\Theta_{jr} = \left\{\omega : \sum_{k \in U_{jr}} |\widehat{b}_{jk} - b_{jk}|^2 \geq 0.25\mu^2\ (n^*)^{-1} 2^{2\nu j} \ln(n^*) j^\lambda. \right\}$$

To construct the upper bounds for R_3 and R_4, note that simple algebra yields $R_3 \le (R_{31} + R_{32})$, $\quad R_4 \le (R_{41} + R_{42})$, where

$$R_{31} = \sum_{j=j_0}^{J-1} \sum_{r \in A_j} \sum_{k \in U_{jr}} \mathbb{E}\left[(\widehat{b}_{jk} - b_{jk})^2 \, \mathbb{I}(\Theta_{jr}) \right], \quad R_{41} = \sum_{j=j_0}^{J-1} \sum_{r \in A_j} \sum_{k \in U_{jr}} \mathbb{E}\left[b_{jk}^2 \, \mathbb{I}(\Theta_{jr}) \right],$$

$$R_{32} = \sum_{j=j_0}^{J-1} \sum_{r \in A_j} \sum_{k \in U_{jr}} \mathbb{E}\left[(\widehat{b}_{jk} - b_{jk})^2 \, \mathbb{I}\big(B_{jr} > 0.25\mu^2 \, (n^*)^{-1} 2^{2\nu j} \ln(n^*) j^{\lambda}\big) \right],$$

$$R_{42} = \sum_{j=j_0}^{J-1} \sum_{r \in A_j} \sum_{k \in U_{jr}} \mathbb{E}\left[b_{jk}^2 \, \mathbb{I}\big(B_{jr} < 2.5\mu^2 \, (n^*)^{-1} 2^{2\nu j} \ln(n^*) j^{\lambda}\big) \right].$$

Then, by (9.9), Lemmas 1 and 2, and the Cauchy-Schwartz inequality, we derive

$$\begin{aligned} R_{31} + R_{41} &= O\left(\sum_{j=j_0}^{J-1} \sum_{r \in A_j} \sum_{k \in U_{jr}} \left[\sqrt{\mathbb{E}(\widehat{b}_{jk} - b_{jk})^4} + b_{jk}^2 \right] \sqrt{\mathbb{P}(\Theta_{jr})} \right) \\ &= O\left(\sum_{j=j_0}^{J-1} [\sqrt{n} \, (\ln n)^{3\lambda/2} \, (n^*)^{-\frac{3}{2(2\nu+1)}} + 2^{-2js^*}] \, n^{-\frac{3}{2}} \right) = O((n^*)^{-1}), \end{aligned}$$

provided μ satisfies (3.19). Hence,

$$\Delta_1 = R_{31} + R_{41} = O\big((n^*)^{-1}\big). \tag{9.13}$$

Now, consider

$$\Delta_2 = R_{32} + R_{42}. \tag{9.14}$$

First, let us study the dense case, i.e., when $\nu(2-p) < ps^*$. Let j_1 be such that

$$2^{j_1} = (n^*)^{\frac{1}{2s+2\nu+1}} (\ln n)^{\frac{(2/p-1)_+ - \lambda}{2\nu+2s+1}}. \tag{9.15}$$

Then, Δ_2 can be partitioned as $\Delta_2 = \Delta_{21} + \Delta_{22}$, where the first component is calculated over the set of indices $j_0 \le j \le j_1$ and the second component over $j_1 + 1 \le j \le J-1$. Hence, using (2.8) and Lemma 1, and taking into account that the cardinality of A_j is $|A_j| = 2^j / \ln n$, we obtain

$$\Delta_{21} = O\left(\sum_{j=j_0}^{j_1} \left[\frac{2^{(2\nu+1)j} j^{\lambda}}{n^*} + \sum_{r \in A_j} \frac{2^{2\nu j} \ln(n^*) j^{\lambda}}{n^*} \right] \right) = O\left(\left[\frac{(\ln n)^{\lambda}}{n^*} \right]^{\frac{2s}{2s+2\nu+1}} (\ln n)^{\varrho} \right), \tag{9.16}$$

where ϱ is defined in (3.14). To obtain an expression for Δ_{22}, note that for $p \ge 2$, by (3.10) and (9.9), we have

$$\Delta_{22} = O\Big(\sum_{j=j_1+1}^{J-1} \sum_{r \in A_j} B_{jr} \Big) = O\Big(\sum_{j=j_1+1}^{J-1} 2^{-2js} \Big) = O\Big((n^*)^{-\frac{2s}{2s+2\nu+1}} (\ln n)^{\frac{2s\lambda}{2s+2\nu+1}} \Big). \tag{9.17}$$

If $1 \le p < 2$, then $B_{jr}^{p/2} = \left(\sum_{k \in U_{jr}} b_{jk}^2 \right)^{p/2} \le \sum_{k \in U_{jr}} |b_{jk}|^p$, so that by Lemma 1, and since

$\nu(2-p) < ps^*$, we obtain

$$\begin{aligned}
\Delta_{22} &= O\left(\sum_{j=j_1+1}^{J-1} \sum_{r \in A_j} \left[\left((n^*)^{-1} 2^{2\nu j} j^{\lambda} \ln n \right)^{1-p/2} \; B_{jr}^{p/2} \right] \right) \\
&= O\left(\sum_{j=j_1+1}^{J-1} \left((n^*)^{-1} 2^{2\nu j} j^{\lambda} \ln n \right)^{1-p/2} 2^{-pjs^*} \right) = O\left((n^*)^{-\frac{2s}{2s+2\nu+1}} (\ln n)^{\frac{2s\lambda}{2s+2\nu+1}+\varrho} \right).
\end{aligned} \tag{9.18}$$

Let us now study the sparse case when $\nu(2-p) > ps^*$. Let j_1 be defined by $2^{j_1} = (n^*)^{\frac{1}{2s+2\nu+1}} (\ln n)^{\frac{-\lambda}{2\nu+2s+1}}$. Hence, if $B_{jr} \geq 0.25\mu^2 \; (n^*)^{-1} 2^{2\nu j} \ln(n^*) j^{\lambda}$, then $B_{jr} \leq \sum_{k=0}^{2^j-1} b_{jk}^2 \leq c^* 2^{-2js^*}$ (see (9.9)) implies that $j \leq j_2$ where j_2 is such that $2^{j_2} = C[n^*/(\ln n)^{1+\lambda}]^{1/(2s^*+2\nu)}$, where C depends on μ and c^* only. Again, partition $\Delta_2 = \Delta_{21} + \Delta_{22}$, where the first component is calculated over $j_0 \leq j \leq j_2$ and the second component over $j_2+1 \leq j \leq J-1$. Then, using similar arguments to that in (9.18), and taking into account that $\nu(2-p) > ps^*$, we derive

$$\begin{aligned}
\Delta_{21} &= O\left(\sum_{j=j_0}^{j_2} \left[(n^*)^{-1} 2^{2\nu j} j^{\lambda} \ln n \right]^{1-p/2} \sum_{r \in A_j} \sum_{k \in U_{jr}} |b_{jk}|^p \right) \\
&= O\left(\sum_{j=j_0}^{j_2} \left[(n^*)^{-1} 2^{2\nu j} j^{\lambda} \ln n \right]^{1-p/2} 2^{-pjs^*} \right) = O\left([(n^*)^{-1} (\ln n)^{1+\lambda}]^{\frac{2s^*}{2s^*+2\nu}} \right).
\end{aligned} \tag{9.19}$$

To obtain an upper bound for Δ_{22}, recall (9.14) and keep in mind that the portion of R_{32} corresponding to $j_2+1 \leq j \leq J-1$ is just zero. Hence, by (9.9), we obtain

$$\Delta_{22} = O\left(\sum_{j=j_2+1}^{J-1} \sum_{k=0}^{2^j-1} b_{jk}^2 \right) = O\left(\sum_{j=j_2+1}^{J-1} 2^{-2js^*} \right) = O\left([(n^*)^{-1} (\ln n)^{1+\lambda}]^{\frac{2s^*}{2s^*+2\nu}} \right).$$

Now, in order to complete the proof, we just need to study the case when $\nu(2-p) = ps^*$. In this situation, we have $2s/(2s+2\nu+1) = 2s^*/(2s^*+2\nu) = 1-p/2$ and $2\nu j(1-p/2) = pjs^*$. Recalling (3.1) and noting that $s^* \leq s'$, we obtain $\sum_{j=j_0}^{J-1} \left(2^{pjs^*} \; \sum_{k=0}^{2^j-1} |b_{jk}|^p \right)^{q/p} \leq A^q$. Then, we repeat the calculations in (9.19) for all indices $j_0 \leq j \leq J-1$. If $1 \leq p < q$, then, by Hölder's inequality, we obtain

$$\begin{aligned}
\Delta_2 &= O\Bigg(\left((n^*)^{-1} (\ln n)^{1+\lambda} \right)^{1-p/2} \; (\ln n)^{1-p/q} \; \left[\sum_{j=j_0}^{J-1} \left(2^{pjs^*} \; \sum_{k=0}^{2^j-1} |b_{jk}|^p \right)^{q/p} \right]^{p/q} \Bigg) \\
&= O\Bigg(\left((n^*)^{-1} (\ln n)^{1+\lambda} \right)^{\frac{2s^*}{2s^*+2\nu}} \; (\ln n)^{1-p/q} \Bigg).
\end{aligned} \tag{9.20}$$

If $1 \leq q \leq p$, then, by the inclusion $B_{p,q}^s(A) \subset B_{p,p}^s(A)$, we obtain

$$\Delta_2 = O\Bigg(\sum_{j=j_0}^{J-1} \left((\ln n)^{1+\lambda}/n^* \right)^{1-p/2} \; 2^{pjs^*} \sum_{k=0}^{2^j-1} |b_{jk}|^p \Bigg) = O\Bigg(\left((\ln n)^{1+\lambda}/n^* \right)^{\frac{2s^*}{2s^*+2\nu}} \Bigg). \tag{9.21}$$

By combining (9.11)–(9.12), (9.13), (9.16)–(9.21), we complete the proof of Theorem 2. □

Proof of Theorem 3. The first part of the theorem is identical to Proposition 1 of Pensky and Sapatinas (2009). The second part can be proved by contradiction. Assume that, assumptions (4.2) and (4.3) hold but condition (4.4) does not take place. It follows from (4.2) and (4.3) that

$$|g_m(u_*)|^2 \le K\,|m|^{-2\nu_1}\exp(-\alpha_1|m|^{\beta_1}), \quad \nu_1>0 \quad \text{if} \quad \alpha_1=0,$$
$$|g_m(u^*)|^2 \ge K\,|m|^{-2\nu_2}\exp(-\alpha_2|m|^{\beta_2}), \quad \nu_2>0 \quad \text{if} \quad \alpha_2=0,$$

Observe that condition (4.4) of Theorem 3 can be violated only in one of the following ways: $\alpha_1=\alpha_2=0$ but $\nu_2<\nu_1$, or $\alpha_1>0$ but $\alpha_2=0$, or $\alpha_1>0$ and $\alpha_2>0$ but $\beta_2<\beta_1$.

Applying Theorem 1 with $M=1$, $\varepsilon_n=1$ and $u_1=u_*$, we arrive at, as $n\to\infty$,

$$R_n^d(B^s_{p,q}(A),u_*,1) \ge \begin{cases} Cn^{-\frac{2s}{2s+2\nu_1+1}}, & \text{if} \quad \alpha_1=0,\ \nu_1(2-p)<ps^*, \\ C\left(\frac{\ln n}{n}\right)^{\frac{2s^*}{2s^*+2\nu_1}}, & \text{if} \quad \alpha_1=0,\ \nu_1(2-p)\ge ps^*, \\ C(\ln n)^{-\frac{2s^*}{\beta_1}}, & \text{if} \quad \alpha_1>0. \end{cases}$$

On the other hand, applying Theorem 2 with $M=1$, $\varepsilon_n=1$ and $u_1=u^*$, we derive that, as $n\to\infty$,

$$\sup_{f\in B^s_{p,q}(A)} \mathbb{E}\|\hat f_n^d - f\|^2 \le \begin{cases} Cn^{-\frac{2s}{2s+2\nu_2+1}}\ (\ln n)^{\varrho}, & \text{if} \quad \alpha_2=0,\ \nu_2(2-p)<ps^*, \\ C\left(\frac{\ln n}{n}\right)^{\frac{2s^*}{2s^*+2\nu_2}}\ (\ln n)^{\varrho}, & \text{if} \quad \alpha_2=0,\ \nu_2(2-p)\ge ps^*, \\ C(\ln n)^{-\frac{2s^*}{\beta_2}}, & \text{if} \quad \alpha_2>0, \end{cases}$$

where ρ is given by formula (3.14) with $\nu=\nu_2$. Now, to complete the proof just note that if $\alpha_1=\alpha_2=0$ but $\nu_2<\nu_1$ or $\alpha_1>0$ but $\alpha_2=0$ or $\alpha_1\alpha_2>0$ but $\beta_2<\beta_1$, then the asymptotical minimax lower bounds for the L^2-risk at the point $u=u_*$ are higher than the corresponding upper bounds at the point $u=u^*$. Hence, in this case, the convergence rates cannot be independent of the choice of M and the selection of points $\underline{u}$, arriving at the required contradiction. □

Proof of Theorem 4. Note that the first inequality in formula (5.1) as well as relations (5.2) and (5.3) between upper bounds in discrete and continuous cases follow directly from Theorems 1 and 2 and from inequalities $\tau_1^d(m,u^*,1)\ge K\tau_1^c(m)$ and $\tau_1^d(m,u^*,1)\ge K\tau_1^d(m,\underline{u},M)$. Hence, one only needs to prove second asymptotic relation in formula (5.1).

Let $R_n^d(B^s_{p,q}(A),\underline{u},M)$ be the minimax L^2-risk for fixed values of $\underline{u}$ and M, defined by formula (3.3), and let

$$H(\underline{u},M,j)=2^{-j}\,\gamma_j^2\sum_{m\in C_j}\tau_1^d(m,\underline{u},M).$$

From the proof of Theorem 1 of Pensky & Sapatinas (2009), it follows that, in the sparse case (when $\nu(2-p)\ge ps^*$), one has $R_n^d(B^s_{p,q}(A),\underline{u},M)\ge C2^{-2j_ns^*}$, as $n\to\infty$, where $j_n\equiv j_n(\underline{u},M)$ is such that $nH(\underline{u},M,j_n)/j_n=C$. Similarly, in the dense case (when $\nu(2-p)<ps^*$), one has $R_n^d(B^s_{p,q}(A),\underline{u},M)\ge C2^{-2j_ns}$, as $n\to\infty$, where $j_n\equiv j_n(\underline{u},M)$ is such that $nH(\underline{u},M,j_n)=C$.

Consider now two different values of M, say M_1 and M_2, and the corresponding sets of $\underline{u}$'s, say $\underline{u}_1$ and $\underline{u}_2$. If $\tau_1^d(m,\underline{u}_1,M_1)<\tau_1^d(m,\underline{u}_2,M_2)$ for any $m\in C_j$, then $H(\underline{u}_1,M_1,j)<H(\underline{u}_2,M_2,j)$. Observe that, for fixed M and $\underline{u}$, both $H(\underline{u},M,j)$ and $H(\underline{u},M,j)/j$ are decreasing functions of j. Hence, if $j_{n1}=j_n(\underline{u}_1,M_1)$ and $j_{n2}=j_n(\underline{u}_2,M_2)$ are the values of j_n corresponding to $(\underline{u}_1,M_1)$ and $(\underline{u}_2,M_2)$, respectively, then $j_{n1}\le j_{n2}$. To show that this is true in the dense case, observe that the opposite, $j_{n1}>j_{n2}$, implies $Cn^{-1}=H(\underline{u}_1,M_1,j_{n1})<H(\underline{u}_1,M_1,j_{n2})<H(\underline{u}_2,M_2,j_{n2})$, so that

j_{n2} cannot be the solution of equation $H(\underline{u}_2, M_2, j_{n2}) = Cn^{-1}$ and $j_{n1} > j_{n2}$ cannot be true. In the sparse case, one just needs to replace $H(\underline{u}, M, j)$ by $H(\underline{u}, M, j)/j$.

Now, it follows immediately that in both sparse and dense cases, $R_n^d(B_{p,q}^s(A), \underline{u}_1, M_1) > R_n^d(B_{p,q}^s(A), \underline{u}_2, M_2)$. Therefore, the infimum of $R_n^d(B_{p,q}^s(A), \underline{u}, M)$ is attained at $\tilde{M}$ and $\tilde{\underline{u}}$ such that $\tau_1^d(m, \tilde{\underline{u}}, \tilde{M}) = \sup_{\underline{u}, M} \tau_1^d(m, \underline{u}, M)$. Since, for any choice of M and any selection of points $\underline{u}$, one has $\tau_1^d(m, \underline{u}, M) \leq K\tau_1^d(m; u^*, 1)$, the validity of the theorem follows from Theorem 1 in Pensky & Sapatinas (2009). $\square$

Proof of Lemma 3. Recall that $\tau_1^c(m) = \int_a^b |g_m(u)|^2\,du$ and $\tau_1^d(m, u^*, 1) = |g_m(u^*)|^2$. Observe that since $\nu(\cdot)$, $\alpha(\cdot)$ and $\beta(\cdot)$ are continuous functions on the interval $U = [a, b]$, then there exist $\nu_1 \leq \nu_2$, $\alpha_1 \leq \alpha_2$ and $\beta_1 \leq \beta_2$ such that $\nu_1 \leq \nu(u) \leq \nu_2$, $\alpha_1 \leq \alpha(u) \leq \alpha_2$ and $\beta_1 \leq \beta(u) \leq \beta_2$, $u \in [a, b]$. Moreover, in the inequalities above, either $\alpha_1 = \alpha_2 = 0$ and $\nu_1 > 0$, or $\alpha_1 > 0$ and $\beta_1 > 0$. Consider the cases when a) $\alpha(u) \equiv 0$ and b) $\alpha(u) > 0$, $\beta(u) \neq$ const.

Case 1: $\alpha(u) \equiv 0$. Then $|g_m(u)|^2 \asymp |m|^{-2\nu(u)}$, so that $|g_m(u)|^2 \leq K|g_m(u^*)|^2$ and $|g_m(u^*)|^2 \asymp |m|^{-2\nu(u^*)}$. Hence, in the discrete case, the asymptotical minimax lower bounds in (5.17) and the asymptotical minimax upper bounds in (5.18), for the L^2-risk, follow directly from Theorems 1 and 2, respectively.

In order to complete the proof, we need to obtain the asymptotical minimax lower and upper bounds for the L^2-risk in the continuous model. For this purpose, observe that, under conditions (5.8) and (5.10), one has (see, e.g., Bender & Orzag (1978), pp. 266–267)

$$\tau_1^c(m) \asymp \int_a^b \exp(-2|\ln m|\nu(u))\ du \asymp |m|^{-2\nu(u^*)}(\ln|m|)^{-1/k}, \tag{9.22}$$

so that Theorems 1 and 2 yield, respectively, the asymptotical minimax lower bounds in (5.11) and the asymptotical minimax upper bounds in (5.12), for the L^2-risk.

Case 2: $\alpha(u) > 0$ and $\beta(u) \neq$ const. In this case, $\beta(u^*) = \beta_1$. Therefore, one derives $K|m|^{-2\nu_2}\exp(-\alpha_2|m|^{\beta_1}) \leq |g_m(u^*)|^2 \leq K|m|^{-2\nu_1}\exp(-\alpha_1|m|^{\beta_1})$ Hence, the asymptotical minimax lower bounds in (5.17) and the asymptotical minimax upper bounds in (5.18), for the L^2-risk, follow directly from Theorems 1 and 2, respectively.

To obtain the asymptotical minimax lower and upper bounds for the L^2-risk in the continuous model, note that

$$\tau_1^c(m) \leq K|m|^{-2\nu_1}\int_a^b \exp\left(-\alpha(u)|m|^{\beta(u)}\right)\,du \leq C_3(b-a)|m|^{-2\nu_1}\ \exp(-\alpha_1|m|^{\beta_1}). \tag{9.23}$$

On the other hand,

$$\tau_1^c(m) \geq K|m|^{-2\nu_2}\ \int_a^b \exp\left(-\alpha(u)|m|^{\beta(u)}\right)\,du. \tag{9.24}$$

Since $\beta(\cdot)$ is a continuously differentiable function in some neighborhood of u^*, $|u-u^*| < d$, we have $\beta(u) \leq \beta(u^*) + \beta_*|u-u^*|$, where $\beta^* = \max_{|u-u^*|<d}|\beta'(u)|$. Therefore, using the inequality $e^z < 1+3z$ for $0 < z < 1$, we obtain $|m|^{\beta(u)} \leq |m|^{\beta_1}\exp(\beta_*|u-u^*|\ln|m|) \leq |m|^{\beta_1}(1+3\beta_*|u-u^*|\ln|m|)$ for $|u-u^*| < \ln|m|/(3\beta_*)$. Denote $\Omega_m(u^*) = \left\{u \in U:\ |u-u^*| < |m|^{-(\beta_1+1)}\right\}$. Then,

$$\begin{aligned}\int_a^b \exp\left(-\alpha(u)|m|^{\beta(u)}\right)\,du &\geq e^{-\alpha_2|m|^{\beta_1}}\ \int_{\Omega_m(u^*)} \exp\left(-\alpha_2|m|^{\beta_1}\,3\beta_*|u-u^*|\ln|m|\right)\,du \\ &\geq e^{-1}\,|m|^{-(\beta_1+1)}\exp(-\alpha_2|m|^{\beta_1}),\end{aligned} \tag{9.25}$$

since $3\beta_*\alpha_2|m|^{-1}\ln|m| < 1$ for $|m|$ large enough. Combining (9.23)– (9.25), we derive that

$$Ke^{-1}\,|m|^{-(2\nu_2+\beta_1+1)}\,\exp\left(-\alpha_2|m|^{\beta_1}\right) \le \tau_1^c(m) \le K(b-a)|m|^{-2\nu_1}\,\exp\left(-\alpha_1|m|^{\beta_1}\right), \tag{9.26}$$

so that Theorems 1 and 2 yield, respectively, the asymptotical minimax lower bounds in (5.11) and the asymptotical minimax upper bounds in (5.12), for the L^2-risk. □

Proof of Theorem 5. First consider the case when $\alpha(u) \equiv 0$. From (5.8), it follows that $\tau_1^d(m,\underline{u},M_n) \le K|m|^{-2\nu(u^*)}$, so that $\varepsilon_n \le K(\ln|m|)^{\lambda_1}$. Since, in this case, $\ln|m| \asymp \ln n$ and $\ln|m| > 1$ as $n\to\infty$, one has $\varepsilon_n = O((\ln n)^{\lambda_3})$, where $\lambda_3 = \max(\lambda_1, 0)$. The latter, in combination with (5.14), implies that condition (3.10) holds and, moreover, that $Cn(\ln n)^{-\lambda_2} \le n^* \le Cn(\ln n)^{\lambda_3}$. Then, Theorems 1 and 2 imply that under conditions (5.8), (5.13) and (5.14), one has $R_n^d(B_{p,q}^s(A),\underline{u},M) \ge R_n^d(B_{p,q}^s(A))$, where $R_n^d(B_{p,q}^s(A))$ is given by expression (5.17) and that, as $n\to\infty$,

$$\sup_{f\in B_{p,q}^s(A)} \mathbb{E}\|\hat f_n^d - f\|^2 \le \begin{cases} C\left(n^{-1}(\ln n)^{\lambda_1+\lambda_2}\right)^{-\frac{2s}{2s+2\nu(u^*)+1}}\,(\ln n)^{\varrho}, & \text{if}\quad \nu(u^*)(2-p) < ps^*,\\ C\Big(n^{-1}(\ln n)^{1+\lambda_1+\lambda_2}\Big)^{\frac{2s^*}{2s^*+2\nu(u^*)}}\,(\ln n)^{\varrho}, & \text{if}\quad \alpha(u)=0,\ \nu(u^*)(2-p)\ge ps^*, \end{cases}$$

where ρ is defined in (3.14). If, moreover, (5.15) holds, then Theorem 1 yields, as $n\to\infty$,

$$R_n^d(B_{p,q}^s(A),\underline{u},M) \ge \begin{cases} C\left(n^{-1}(\ln n)^{\lambda_1+\lambda_2}\right)^{-\frac{2s}{2s+2\nu(u^*)+1}}, & \text{if}\quad \nu(u^*)(2-p) < ps^*,\\ C\Big(n^{-1}(\ln n)^{1+\lambda_1+\lambda_2}\Big)^{\frac{2s^*}{2s^*+2\nu(u^*)}}, & \text{if}\quad \nu(u^*)(2-p)\ge ps^*. \end{cases}$$

To complete the proof of this part, compare the above upper and lower bounds with (5.11) and (5.12).

Now, let $\alpha(u) > 0$. Then, due to assumption (3.10) one has $\ln n^* \asymp \ln n$. Under condition (5.8), by Theorem 1, $R_n^d(B_{p,q}^s(A),\underline{u},M) \ge R_n^d(B_{p,q}^s(A)) \ge C(\ln n)^{-2s^*/\beta(u^*)}$, as $n\to\infty$. Also, by Theorem 2, $\sup_{f\in B_{p,q}^s(A)} \mathbb{E}\|\hat f_n^d - f\|^2 \le C(\ln n)^{-2s^*/\beta(u^*)}$, as $n\to\infty$. To complete the proof, compare the above lower and upper bounds for the L^2-risks with the corresponding bounds in (5.11) and (5.12). □

Proof of Theorem 6. Note that conditions (5.19), (5.22) and Theorem 1 imply that, as $n\to\infty$,

$$R_n^d(B_{p,q}^s(A),\underline{u},M) \ge \begin{cases} C(n^*)^{-\frac{2s}{2s+2\nu+1}}\,(\ln n^*)^{\frac{2s\lambda}{2s+2\nu+1}}, & \text{if}\quad \alpha=0,\ \nu(2-p)<ps^*,\\ C\left(\frac{\ln n^*}{n^*}\right)^{\frac{2s^*}{2s^*+2\nu}}(\ln n^*)^{\frac{2s^*\lambda}{2s^*+2\nu}}, & \text{if}\quad \alpha=0,\ \nu(2-p)\ge ps^*,\\ C(\ln n^*)^{-\frac{2s^*}{\beta}}, & \text{if}\quad \alpha>0. \end{cases} \tag{9.27}$$

Denote the ratio between the upper bound for the L^2-risk (5.18) in the continuous model and the lower bound (9.27) by $\Delta_n = \sup_{f\in B_{p,q}^s(A)} \mathbb{E}\|\hat f_n^{d*} - f\|^2 \Big/ R_n^d(B_{p,q}^s(A),\underline{u},M)$ and observe that the convergence rates in the discrete model are inferior to the convergence rates in the continuous model if $\lim_{n\to\infty}(\ln n)^{h\,\mathbb{I}(\alpha(u)\equiv 0)}\ \Delta_n = 0$ for any $h>0$.

Let $\alpha(u)\equiv 0$ and consider the case when $\nu(2-p) < ps^*$. Then, taking into account that under condition (3.10) one has $\ln n^* \asymp \ln n$, we obtain

$$\begin{aligned}\lim_{n\to\infty}\Delta_n\,(\ln n)^h &= O\left(\lim_{n\to\infty} n^{-\frac{2s}{2s+2\nu(u^*)+1}}\,(\ln n)^{\varrho+h}\,(n^*)^{\frac{2s}{2s+2\nu+1}}\,(\ln n^*)^{-\frac{2s\lambda}{2s+2\nu+1}}\right)\\ &= O\left(\lim_{n\to\infty}\left[(\ln n)^{\rho+h-\frac{2s\lambda}{2s+2\nu+1}}\,n^{-\left(\frac{2s}{2s+2\nu(u^*)+1}-\frac{2s}{2s+2\nu+1}(1+\varepsilon_0)\right)}\right]\right).\end{aligned}$$

Now, if $\nu > \nu(u^*)$, then it is easy to see that under condition (5.20) we have $\lim_{n\to\infty} \Delta_n (\ln n)^h = 0$ for any h, and the convergence rates in the discrete model are inferior in this case. If $\nu = \nu(u^*)$, then $\lim_{n\to\infty} \Delta_n \ (\ln n)^h = O\left(\lim_{n\to\infty} \left[(\ln n)^{\rho+h-\frac{2s\lambda}{2s+2\nu+1}} \ (\varepsilon_n)^{\frac{2s}{2s+2\nu+1}}\right]\right) = 0$ if condition (5.21) holds. The sparse case when $\nu(2-p) < ps^*$ can be treated in a similar manner.

Now, consider the case when $\alpha(u) > 0$. One has $\lim_{n\to\infty} \Delta_n = \lim_{n\to\infty} (\ln n)^{-\frac{2s^*}{\beta(u^*)}} (\ln n^*)^{\frac{2s^*}{\beta}} = (1+\varepsilon_0)^{\frac{2s^*}{\beta}} \ \lim_{n\to\infty} (\ln n)^{-2s^*\left(\frac{1}{\beta(u^*)}-\frac{1}{\beta}\right)}$, and it is easy to see that under each set of conditions in (5.23), $\lim_{n\to\infty} \Delta_n = 0$. □

Proof of Corollary 1. Note that if $M = M_n$ is finite, then for $\alpha(u) \equiv 0$ one has $\tau_1^d(m, \underline{u}, M_n) \asymp |m|^{-2\nu}$, where $\nu = \min(\nu(u_1), \nu(u_2), \ldots, \nu(u_M))$. If $\alpha(u) > 0$, then denote $l_0 = \arg\min_l \beta(u_l)$, $\beta = \beta(u_{l_0})$, $\nu_0 = \nu(u_{l_0})$ and $\alpha_0 = \alpha(u_{l_0})$. In this case, $\tau_1^d(m, \underline{u}, M_n) \asymp |m|^{-2\nu_0} \exp\left(-\alpha_0 |m|^\beta\right)$ and, hence, the validity of the corollary follows from Theorems 5 and 6. □

Proof of Corollary 2. Note that $\tau_1^d(m, \underline{u}, M_n) \geq K(\ln n)^{-\lambda^*} |m|^{-2\nu(u^*)}$ and, hence, the validity of the corollary follows Theorem 5. □

Proof of Corollary 3. Note that, for u_l such that $\beta(u_l) = \beta(u^*)$ one has $\tau_1^d(m, \underline{u}, M_n) \geq K\ n^{-\tau} |m|^{-2\nu(u_l)} \exp\left(-\alpha(u_l)|m|^{\beta(u^*)}\right)$. Then, the validity of the corollary follows from Theorem 5. □

Proof of Theorem 7.

First, consider the case when $\alpha(u) \equiv 0$. Denote $v(x) = \nu(S(x))$, $x^* = q(u^*)$ and let l^* be the index of a point closest to x^*, i.e., $l^* = \arg\min |x^* - (l-1+d)/M|$. Note that $v(x^*) = \nu(u^*)$ and the function $v(x)$ is continuously differentiable with $|v'(x)| \leq v_0$ for some $v_0 > 0$. Note that $\tau_1^d(m, \underline{u}, M_n) \leq K|m|^{-2\nu(u^*)}$, so if we show that under condition (5.25) we have

$$\tau_1^d(m, \underline{u}, M_n) \geq K|m|^{-2\nu(u^*)} (\ln n)^{-\lambda} \tag{9.28}$$

for some constant $\lambda \in \mathbb{R}$, then the validity of the theorem will follow from Theorem 5. In order to prove (9.28), note that

$$\begin{aligned}
\tau_1^d(m, \underline{u}, M) &\asymp \frac{1}{M} \sum_{l=1}^M |m|^{-2v\left(\frac{l-1+d}{M}\right)} \\
&\asymp \frac{1}{M} \sum_{l=1}^M |m|^{-2\left[v\left(\frac{l-1+d}{M}\right) - v\left(\frac{l^*-1+d}{M}\right)\right] - 2\left[v\left(\frac{l^*-1+d}{M}\right) - v(x^*)\right] - 2\nu(u^*)} \\
&\geq \frac{K}{M} |m|^{-2\nu(u^*)} \sum_{l=1}^M |m|^{-\frac{2v_0|l-l^*|}{M} - \frac{v_0^*}{M}} \geq \frac{K}{M} |m|^{-2\nu(u^*)} \sum_{k=0}^{M/2-1} |m|^{-\frac{2v_0 k}{M} - \frac{v_0^*}{M}},
\end{aligned}$$

where $v_0^* = v_0 \mathbb{I}(x^* \neq (l^*-1+d)/M)$. Now, recall the following statement from Calculus: if $u(z)$, $z \geq 0$, is a continuous, positive, monotonically decreasing function, then

$$\sum_{k=0}^{M/2-1} u(k) \geq \max\left(\int_0^{M/2} u(x)dx, u(0) + \int_1^{M/2} u(x)dx\right) \geq \frac{1}{2} \ \left(u(0) + \int_0^{M/2} u(x)dx\right). \tag{9.29}$$

Applying (9.29) with $u(x) = |m|^{-2\nu(u^*) - \frac{v_0^*}{M}} |m|^{-\frac{2v_0 x}{M}}$, and taking into account that $\int_0^{M/2} u(x)dx \asymp |m|^{-2\nu(u^*) - \frac{v_0^*}{M}} \ M^{-1} \ln|m|$, we obtain

$$\tau_1^d(m, \underline{u}, M) \geq K\, |m|^{-2\nu(u^*)} (\ln|m|)^{-1} \ \left(1 + M^{-1} \ln|m|\right) \times \exp\left(-v_0^* M^{-1} \ln|m|\right).$$

Now, recall that $\ln|m| \asymp \ln n$ in this case and note that under the first assumption in (5.25), $\tau_1^d(m, \underline{u}, M)$ satisfies condition (5.13) of Theorem 5 with $\lambda_1 = -1$ and $\varepsilon_n = 1$. Hence, the convergence rates in the discrete and the continuous models almost coincide.

Now, consider the case when $\alpha(u) > 0$. Denote $v(x) = \beta(S(x))$ and let, as before, $x^* = q(u^*)$ and $l^* = \arg\min|x^* - (l-1+d)/M|$. Note that $v(x^*) = \beta(u^*)$ and that the function $v(x)$ is continuously differentiable with $|v'(x)| \leq v_0$ for some constant $v_0 > 0$. Denote, as before, $v_0^* = v_0 \mathbb{I}(x^* \neq (l^*-1+d)/M)$. Note that $\tau_1^d(m, \underline{u}, M) \geq M^{-1} K\,|m|^{-2\nu_1} \sum_{l=1}^M \exp\left(-\alpha_1 |m|^{\beta(u_l)}\right)$, where $\nu_1 = \max \nu(u)$, $\alpha_1 = \max \alpha(u)$, $u \in U$, and, in order to prove the statement, we need to construct a lower bound for $S(m,M) = \sum_{l=1}^M \exp\left(-\alpha_1 |m|^{\beta(u_l)}\right)$. Similarly to the polynomial case, we obtain that $S(m,M) \geq K \sum_{k=0}^{M/2-1} \exp\left(-\alpha_1 |m|^{\beta(u^*) + \frac{v_0 k}{M} + \frac{v_0^*}{2M}}\right)$. Denote $\alpha_m = \alpha_1 |m|^{\beta(u^*) + \frac{v_0^*}{2M}}$ and apply the inequality (9.29) with $u(x) = \exp\left(-\alpha_m\,|m|^{\frac{v_0 k}{M}}\right)$. Observe that

$$\int_0^{M/2} u(x)dx \;=\; \frac{M}{v_0^* \ln|m|} \int_1^{|m|^{v_0/2}} z^{-1} \exp(-\alpha_m z) dz \;\geq\; \frac{M}{2 v_0^* \ln|m| \alpha_m} \exp(-\alpha_m),$$

and recall that $\ln|m| \asymp \ln\ln n$. Hence, under the second of conditions in (5.25), as $M \to \infty$ and $|m| \to \infty$, we derive that

$$S(m,M) \;\geq\; K\,M\,(\ln n)^{-1}\,|m|^{-(\beta(u^*)+1)} \exp\left(-\alpha_1 |m|^{\beta(u^*)} \exp[0.5 v_0^* M^{-1} \ln|m|]\right).$$

Note that due to assumption (5.25) and due to $\ln|m| \asymp \ln\ln n$, there exists $\tau_3 > 0$ such that $M_n^{-1} \ln|m| \leq \tau_3$ when n is large enough. Therefore, $\tau_1^d(m, \underline{u}, M) \geq C\,(\ln n)^{-1}\,|m|^{-(2\nu_1 + \beta(u^*)+1)}$ $\exp\left(-\alpha_1 \exp(0.5\, v_0^* \tau_3)\,|m|^{\beta(u^*)}\right)$. Application of Theorem 5 with $\nu = \nu_1 + 0.5\,\beta(u^*) + 0.5$, $\alpha = \alpha_1 \exp(0.5\, v_0^* \tau_3)$ and $\varepsilon_n = (\ln n)^{-1}$ completes the proof of this part of the statement.

To prove the last statement in Theorem 7, recall a simple fact from Calculus: if function $F(x)$, $x \in [0,1]$, is continuously differentiable with $F_0 = \max_x |F'(x)|$, then for any d such that $0 \leq d \leq 1$ one has

$$\left| M^{-1} \sum_{l=1}^M F\left((l-1+d)/M\right) - \int_0^1 F(x)dx \right| \leq 0.5\, M^{-1}\, F_0. \tag{9.30}$$

Let $\alpha(u) \equiv 0$. Note that since $|q'(u)|$ is bounded and separated from zero, one has $\int_0^1 |g_m(S(x))|^2 dx = \int_a^b |g_m(u)|^2 q'(u) du \asymp \tau_1^c(m)$. Therefore, if

$$R(m,n) = \tau_1^d(m, \underline{u}, M_n) - \int_0^1 |g_m(S(x))|^2 dx = o\left(\tau_1^c(m)\right), \tag{9.31}$$

as $n \to \infty$, then $\tau_1^d(m, \underline{u}, M_n) \asymp \tau_1^c(m)$ and the theorem is proved. Applying formula (9.30) to $F(x) = |g_m(S(x))|^2$ and noting that $|S'(x)| \leq s_2$, we obtain

$$R(m,n) \leq 0.5\, s_2\, M_n^{-1} \max_{u \in [a,b]} \left| \frac{d}{du} |g_m(u)|^2 \right| = O\left(M_n^{-1}\,|m|^{-2\nu(u^*)} \ln|m|\right).$$

Comparing the last expression with $\tau_1^c(m)$ given by formula (9.22), we confirm that condition (9.31) holds and theorem is valid in this case. □

Proof of Lemma 4. Recall that $0 < a < b < \infty$, $\beta_1 \leq \beta(u) \leq \beta_2$, $u \in [a,b]$, for some $0 < \beta_1 \leq \beta_2 < \infty$, and $u_l = a + (b-a)l/M$, $l = 1, 2, \ldots, M$. Consider first the case when $a \in \mathbb{N}$.

Then, $4\pi^2 m^2 M\, \tau_1^d(m, \underline{u}, M) = \sum_{l=1}^M \beta^2(u)\sin^2(2\pi m u_l) \geq \beta_1^2 \sum_{l=1}^M \sin^2\big(2\pi m(b-a)l/M\big)$. Using the formula 1.351.1 of Gradshtein & Ryzhik (1980) with $x = 2\pi m(b-a)/M$, $n = M$ and $k = l$, we obtain

$$\tau_1^d(m, \underline{u}, M) \geq \frac{\beta_2^2}{4\pi^2 m^2 M}\left(\frac{M}{2} - \frac{\cos(2(M+1)\pi m(b-a)/M)\sin(2\pi m(b-a))}{2\sin(2\pi m(b-a)/M)}\right). \tag{9.32}$$

Since $2^j \leq |m| < 2\pi/32^j$ and $\ln n \leq 2^j \leq n^{1/3}$, the condition $M_n \geq (32\pi/3)(b-a)n^{1/3}$ guarantees that $|2\pi m(b-a)/M| \leq \pi/2$. In this case, using the inequality $y \leq 2\sin(y)$, $0 \leq y \leq \pi/2$ (see, e.g., Lang (1966), p. 41), we derive

$$2\sin(2\pi m(b-a)/M) \geq 2\pi m(b-a)/M \geq (4\pi^2(b-a)\ln n)/(3M). \tag{9.33}$$

Hence, combining (9.32) and (9.33), for n large enough, we arrive at $\tau_1^d(m, \underline{u}, M) \geq K m^{-2}$.

Consider now the case when $a \notin \mathbb{N}$. A standard trigonometrical identity yields

$$\sum_{l=1}^M \sin^2(2\pi m a + 2\pi m(b-a)l/M) = \frac{1}{2}\Big(M - \sum_{l=1}^M \cos(4\pi m a + 4\pi m(b-a)l/M)\Big). \tag{9.34}$$

Using the formula 1.341.3 of Gradshtein & Ryzhik (1980) with $x = 4\pi m a$, $y = 4\pi m(b-a)/M$, $n = M$ and $k = l$, we derive for $M \geq 4(b-a)|m|$,

$$\begin{aligned}\sum_{l=1}^M \cos(4\pi m a + 4\pi m(b-a)l/M) &= \frac{\cos(4\pi m a + 2\pi m(b-a)(M-1)/M)\sin(2\pi m(b-a))}{\sin(2\pi m(b-a)/M)} \\ &\leq \frac{M}{\pi m(b-a)}.\end{aligned} \tag{9.35}$$

Hence, combining (9.34) and (9.35) in a manner similar to the first part of the proof, for $|m|$ large enough, we arrive at $\tau_1^d(m, \underline{u}, M) \geq K m^{-2}$ which completes the proof. □

Proof of Theorem 8.

The proof follows directly from the discussion of Section 6, by combining Theorem 1, Theorem 2 and Lemma 4, taking $A_j = C_j = \{m : \psi_{mjk} \neq 0\}$, and noting that, for the Meyer wavelets, $C_j \subseteq 2\pi/3[-2^{j+2}, -2^j] \cup [2^j, 2^{j+2}]$ with $|C_j| = 4\pi 2^j$ (see, Johnstone *et.al.* (2004, p. 565). □

Acknowledgments

Marianna Pensky was supported in part by National Science Foundation (NSF), grants DMS-0505133 and DMS-0652524. The authors would like to thank Athanasia Petsa for her help in carrying out the simulation study. Finally, we would like to thank an Associate Editor and two anonymous referees for their suggestions which helped to significantly improve the paper.

References

[1] Abramovich, F. and Silverman, B.W. (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, **85**, 115–129.

[2] Antoniadis, A., Bigot, J. & Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: a comparative simulation study. *Journal of Statistical Software,* **6**, Article **6**.

[3] Bender, C.M. and Orzag, S.A. (1978) *Advanced Mathematical Methods for Scientists and Engineers.* McGraw-Hill, New York.

[4] Brown, L.D., Cai, T., Low, M.G., and Zhang, C.-H. (2002) Asymptotic equivalence theory for nonparametric regression with random design. *Annals of Statistics*, **30**, 688–707.

[5] Brown, L.D.and Low, M.G. (1996) Asymptotic equivalence of nonparametric regression and white noise. *Annals of Statistics*, **24**, 2384–2398.

[6] Casey, S.D. and Walnut, D.F. (1994). Systems of convolution equations, deconvolution, Shannon sampling, and the wavelet and Gabor transforms. *SIAM Review*, **36**, 537–577.

[7] Cavalier, L. and Raimondo, M. (2007). Wavelet deconvolution with noisy eigenvalues. *IEEE Transactions on Signal Processing*, **55**, 2414–2424.

[8] Chesneau, C. (2008). Wavelet estimation via block thresholding: a minimax study under L^p-risk. *Statistica Sinica*, **18**, 1007–1024.

[9] Cirelson, B.S., Ibragimov, I.A. and Sudakov, V.N. (1976). Norm of Gaussian sample function. In *Proceedings of the 3rd Japan-U.S.S.R. Symposium on Probability Theory*, Lecture Notes in Mathematics, Vol. **550**, 20–41. Springer-Verlag, Berlin.

[10] De Canditiis, D. and Pensky, M. (2004). Discussion on the meeting on "Statistical Approaches to Inverse Problems". *Journal of the Royal Statistical Society*, Series B, **66**, 638–640.

[11] De Canditiis, D. and Pensky, M. (2006). Simultaneous wavelet deconvolution in periodic setting. *Scandinavian Journal of Statistics*, **33**, 293–306.

[12] Donoho, D.L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Applied and Computational Harmonic Analysis*, **2**, 101–126.

[13] Donoho, D.L. and Raimondo, M. (2004). Translation invariant deconvolution in a periodic setting. *International Journal of Wavelets, Multiresolution and Information Processing*, **14**, 415–432.

[14] Golubev, G. (2004) The principle of penalized empirical risk in severely ill-posed problems. *Probability Theory and Related Filelds*, **130**, 18–38.

[15] Golubev, G.K. and Khasminskii, R.Z. (1999). A statistical approach to some inverse problems for partial differential equations. *Problems of Information Transmission*, **35**, 136–149.

[16] Gradshtein, I.S. and Ryzhik, I.M. (1980). *Tables of Integrals, Series, and Products.* Academic Press, New York.

[17] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications.* Lecture Notes in Statistics, Vol. **129**, Springer-Verlag, New York.

[18] Harsdorf, S. and Reuter, R. (2000). Stable deconvolution of noisy lidar signals. In *Proceedings of EARSeL-SIG-Workshop LIDAR*, Dresden/FRG, June 16–17.

[19] Johnstone, I.M. (2002). *Function Estimation in Gaussian Noise: Sequence Models.* Unpublished Monograph. (http://www-stat.stanford.edu/∼imj/)

[20] Johnstone, I.M., Kerkyacharian, G., Picard, D. and Raimondo, M. (2004) Wavelet deconvolution in a periodic setting. *Journal of the Royal Statistical Society*, Series B, **66**, 547–573 (with discussion, 627–657).

[21] Johnstone, I.M. and Raimondo, M. (2004). Periodic boxcar deconvolution and Diophantine approximation. *Annals of Statistics*, **32**, 1781–1804.

[22] Kalifa, J. and Mallat, S. (2003). Thresholding estimators for linear inverse problems and deconvolutions. *Annals of Statistics*, **31**, 58–109.

[23] Kerkyacharian, G., Picard, D. and Raimondo, M. (2007). Adaptive boxcar deconvolution on full Lebesgue measure sets. *Statistica Sinica*, **7**, 317–340.

[24] Kolaczyk, E.D. (1994). Wavelet methods for the inversion of certain homogeneous linear operators in the presence of noisy data. *PhD Dissertation*, Department of Statistics, Stanford University, USA.

[25] Lang, S. (1966). *Introduction to Diophantine Approximations.* Springer-Verlag, New York.

[26] Lattes, R. and Lions, J.L. (1967). *Methode de Quasi-Reversibilite et Applications.* Travoux et Recherche Mathematiques, Vol. **15**, Dunod, Paris.

[27] Meyer, Y. (1992). *Wavelets and Operators.* Cambridge University Press, Cambridge.

[28] Müller, H.G. & Stadmüller, U. (1987) Variable bandwidth kernel estimators of regression curves. *Annals of Statistics*, **15**, 182–201.

[29] Neelamani, R., Choi, H. and Baraniuk, R. (2004). Forward: Fourier-wavelet regularized deconvolution for ill-conditioned systems. *IEEE Transactions on Signal Processing*, **52**, 418–433.

[30] Park, Y.J., Dho, S.W and Kong, H.J. (1997). Deconvolution of long-pulse lidar signals with matrix formulation. *Applied Optics*, **36**, 5158–5161.

[31] Pensky, M. and Sapatinas, T. (2009). Functional deconvolution in a periodic case: uniform case. *Annals of Statistics*, **37**, 73–104.

[32] Pensky, M. and Sapatinas, T. (2009a). Diophantine approximation and the problem of estimation of the initial speed of a wave on a finite interval in a stochastic setting. *Preprint.*

[33] Petsa, A. and Sapatinas, T. (2009). Minimax convergence rates under the L^p-risk in the functional deconvolution model. *Statistics and Probability Letters*, **79**, 1568–1576. (Erratum: *Statistics and Probability Letters*, **79**, 1890 (2009).)

[34] Reiss, M. (2008) Asymptotic equivalence for nonparametric regression with multivariate and random design. *Annals of Statistics*, **36**, 1957–1982.

[35] Schmidt, W. (1980). *Diophantine Approximation.* Lecture Notes in Mathematics, Vol. **785**, Springer-Verlag, Berlin.

[36] Strauss, W.A. (1992). *Partial Differential Equations: An Introduction.* John Wiley & Sons, New York.

[37] Willer, T. (2005). Deconvolution in white noise with a random blurring function. *Preprint.* (arXiv:math/0505142v1 [math.ST])